

# AutoMap User's Guide 2007

Kathleen M. Carley, Dave Columbus, Matt DeReno,  
Jana Diesner and Nicole Sebula<sup>1</sup>

October 2006  
CMU-ISRI-07-114

Institute for Software Research  
School of Computer Science  
Carnegie Mellon University  
Pittsburgh, PA 15213

## Abstract

AutoMap is software for computer-assisted Network Text Analysis (NTA). NTA encodes the links among words in a text and constructs a network of the linked words. AutoMap subsumes classical Content Analysis by analyzing the existence, frequencies, and covariance of terms and themes.

---

<sup>1</sup> This work was supported in part by the Office of Naval Research under Contract No. N00014-06-1-0104, ONR, and N00014-06-10921, by the National Science Foundation IGERT in CASOS, the Air Force Office of Sponsored Research with a MURI with George Mason University under Grant No. 600322GRGMASON, and the Army Research Lab under Grant No. DAAD19-01-2-0009. Additional support was provided by the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation, the Office of Naval Research, the Air Force Office of Sponsored Research, the Army Research Lab or the U.S.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>OCT 2006</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2006 to 00-00-2006</b>	
4. TITLE AND SUBTITLE <b>AutoMap User's Guide 2007</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Carnegie Mellon University,School of Computer Science,Pittsburgh,PA,15213</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>178</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

**Key Words:** Semantic Network Analysis, Dynamic Network Analysis, Mental Models, Social Networks, AutoMap

# Table of Contents

AutoMap User's Guide 2007.....	1
Abstract .....	1
Table of Contents .....	1
AutoMap: An Overview .....	11
Network Text Analysis (NTA) .....	11
Semantic Network Analysis.....	11
Social Network Analysis (SNA).....	12
Illustrative problems that people in the DNA area work on - .....	13
Getting Started .....	13
Dynamic Network Analysis.....	14
Where to learn to more: .....	14
Social Network Analysis (SNA).....	14
Where to find out more on SNA.....	15
Automap Graphical User Interface .....	16
Tool Tips.....	17
Load Input.....	17
1. To open a single text.....	17
2. To open a set of texts .....	18
Text Examples .....	18
Our Text Example .....	19
Our Text Example II.....	19
Parts of Speech.....	21
NOTE: Data has to be in txt format with .txt extension.....	21
2. Parts of Speech Tagging.....	22
Parts of Speech.....	23
NOTE: Data has to be in txt format with .txt extension.....	23
2. Parts of Speech Tagging.....	24
Text Pre-Processing .....	24
1. Introduction to Text Pre-Processing in AutoMap.....	24
Tip! All pre-processing techniques in AutoMap are optional.....	25
Named-Entity Recognition .....	25
Stemming.....	25
Collocation Identification .....	25
Tip! Collocations occurring with high frequency are powerful indicators of a pattern of meaning in a text.....	25
Deletion .....	26
Thesauri .....	26
Generalization Thesaurus. ....	26
Meta-Matrix Thesaurus .....	26
Sub-Matrix Selection.....	27
Text Pre-Processing .....	27
2. Hierarchy of Pre-Processing Techniques .....	27

Tip! All pre-processing techniques are optional. ....	27
Text Pre-Processing .....	28
3.1 Named-Entity Recognition .....	28
3.1.1 Example for Named-Entity Recognition .....	29
3.2 Symbol Removal .....	30
Symbol Removal .....	30
3.21 N-gram Identification: Bigrams.....	31
Text PreProcessing.....	31
4. Stemming.....	31
Porter Stemmer: .....	31
Krovetz Stemmer:.....	31
4.1 Example for Stemming (Porter) .....	32
Text PreProcessing.....	33
5. Deletion .....	33
5.1 Open a Delete List .....	34
5.1.1 Small predefined Delete List .....	35
5.1.2 Extensive predefined Delete List.....	35
5.2 Create a Delete List .....	35
5.3. Edit a Delete List .....	36
5.4 Apply a Delete List.....	36
5.4.1 Direct Adjacency .....	37
5.4.2 Rhetorical adjacency.....	37
5.5 Un-Apply a Delete List.....	37
5.6 Save an applied Delete List .....	37
5.7 Save text(s) after application of Delete List .....	37
5.8 Examples for the application of a Delete List .....	37
Text PreProcessing.....	39
6. Generalization Thesaurus .....	39
6.1. Open a Generalization Thesaurus .....	39
6.2. Create a Generalization Thesaurus .....	40
Here is an example for multi-step pre-processing: .....	41
6.3 Edit a Generalization Thesaurus.....	44
6.4.1 Thesaurus content only .....	44
6.4.1.1 Direct Adjacency .....	45
6.4.1.2 Rhetorical Adjacency .....	45
6.5 Un-Apply a Generalization Thesaurus .....	46
6.6 Save an applied Generalization Thesaurus.....	46
6.7 Save text(s) after application of Generalization Thesaurus .....	46
6.8 Example for the building and applying a Generalization Thesaurus.....	46
Applying a thesaurus to text that was pre-processed with a Delete List, direct adjacency: .....	48
Applying a thesaurus to text that was pre-processed with a Delete List, rhetorical adjacency: .....	49
Text PreProcessing.....	51
7. Meta-Matrix Thesaurus .....	51
7.1 Open a Meta-Matrix Thesaurus .....	52

Note: The Meta-Matrix Thesaurus can be edited.....	52
Note: The pre-assigned Meta-Matrix Thesaurus can be edited.....	52
7.2 Edit a Meta-Matrix Thesaurus.....	52
7.3 Build a Meta-Matrix Thesaurus.....	52
7.4 Apply a Meta-Matrix Thesaurus.....	53
7.4.1 Thesaurus content only.....	53
7.4.1.1 Direct Adjacency.....	54
7.4.1.2 Rhetorical Adjacency.....	54
7.5 Un-Apply a Meta-Matrix Thesaurus.....	55
7.6 Save an applied Meta-Matrix Thesaurus.....	55
7.7 Save text(s) after application of Meta-Matrix Thesaurus.....	55
7.8 Example for editing and applying a Meta-Matrix Thesaurus.....	55
Text PreProcessing.....	59
8. Sub-Matrix Selection.....	59
8.1 Select Sub-Matrix Categories!.....	60
Precondition: Sub-Matrix Selection can only be performed if Meta-Matrix Thesaurus was applied. There are 4 ways to select sub matrices:.....	60
8.1.1 If you do not have a sub matrix selection file available.....	60
8.1.2 If you do not have a sub matrix selection file available and want to select the full meta matrix (means all cells in the meta matrix).....	60
8.1.3 If you have a sub matrix selection file available.....	60
8.1.4 If you want to write your own sub matrix selection file and load it into AutoMap:.....	61
8.2 Create or Modify Sub-Matrix Selection.....	61
Tip!.....	61
<input type="checkbox"/> If you wish to analyze any relation between e.g. agents and organizations, you need to select agent/ organization in one row, and organization/ agent in another row. If you wish to analyze any relations among all agents, select agent/agent. ....	61
<input type="checkbox"/> If you wish to analyze any relations among all agents and their relation with all organizations, select agent/agent/organization and in another row organization/agent/agent/. ....	61
8.3 Apply Sub-Matrix Selection.....	61
8.3.1 Direct Adjacency.....	62
8.3.2 Rhetorical Adjacency.....	62
8.4 Un-Apply a Sub-Matrix Selection.....	62
8.5 Save Sub-Matrix Selection.....	62
8.6 Save text(s) after Sub-Matrix Selection.....	63
8.7 Example for Sub-Matrix Selection.....	63
Statement Formation Choices.....	65
1. Analysis Settings.....	65
Overview on the possible Settings:.....	65
1.1 Default Settings.....	66
Anaphora Resolution.....	66
Statement Formation Choices.....	69
1. Analysis Settings.....	69
Overview on the possible Settings:.....	69

1.1 Default Settings .....	70
Semantic Text Analysis .....	70
1. Browse through texts .....	71
1.1 Example for Browse Menu.....	72
2. Concept List.....	73
Tip! The number of unique concepts considers each concept only once, whereas the number of total concepts considers repetitions of concepts. ....	73
2.1 Example for Concept List.....	73
3. Create and refresh Union Concept List .....	74
Note: The number of unique concepts considers each concept only once, whereas the number of total concepts also considers repetitions of concepts. ....	77
3.1 Save Union Concept List.....	77
3.2 Example for Union Concept List.....	78
4. Semantic Network Analysis .....	84
4.1 Example for Semantic Network Analysis .....	85
Note: For more information about the impact of coding choices on map analysis results you might have a look at our publications ( <a href="http://www.casos.cs.cmu.edu/projects/automap/publications.html">http://www.casos.cs.cmu.edu/projects/automap/publications.html</a> ).....	88
Content Analysis .....	88
From content analysis to semantic networks .....	88
I. Open AutoMap .....	88
II. Overview of the AutoMap Graphical User Interface (GUI) .....	89
III. Load multiple text documents into AutoMap.....	89
IV. Carry out a simple content analysis.....	90
V. Save your Union Concept List as a csv file.....	93
VI. View your csv file in Microsoft Excel .....	94
VII. Save and view your results as a binary csv file.....	94
VIII. Carry out a simple Semantic Network Analysis.....	94
Click Open.....	100
IX. Open ORA .....	101
X. Load your files into ORA .....	101
XI. Visualize your meta-matrices .....	101
XII. Run a Semantic Network Report .....	103
SEMANTIC-NETWORK REPORT .....	104
Symmetric difference .....	104
Saved output networks.....	104
Ontological Text Coding.....	105
1. Meta Matrix Text Analysis.....	105
The Meta-Matrix Model used in AutoMap .....	105
Analyzing Text(s) .....	106
1.1 Example for Meta Matrix Text Analysis.....	106
2. Sub Matrix Text Analysis.....	109
2.1 Example for Sub Matrix Text Analysis.....	111
Outputs.....	115
1. Semantic Network .....	115
1.1 Example for Semantic Network Output .....	115

2. Statistics.....	115
2.1 Example for Statistic Output .....	116
3. Additional Outputs .....	116
Output Options.....	116
1. Term Distribution Lists and Matrices.....	117
Points To Consider .....	117
1.1 Types and content of Term Distribution Lists and Matrices .....	118
1.2 Example for Term Distribution List and Matrices .....	119
2. Save Non-Identified concepts.....	120
3. Additional Output Formats.....	121
Note: You can also use the Network Converter to convert data.....	121
3.1 DL:.....	121
Note: Since networks extracted with AutoMap are directed, matrices representing these networks can be rectangular. If a DL file needs to be generated from a rectangular matrix AutoMap by default squares this matrix before converting it into DL format.....	122
3.2 DyNetML: .....	122
Note: Whether to use NTA or SNTA is a "text-philosophical question" and the answer depends upon what the user wants to measure: a textual network or a social network.....	123
3.3 Examples for additional Output Formats.....	123
4. Network Analytic Measures .....	125
Note: Only Network Analytic Measures for directed networks were implemented into AutoMap. The reason for this is that AutoMap outputs are always directed in order to adequately represent the linear structure of texts. ....	125
Note: Texts have a linear structure. Therefore, we only implemented Network Analytic Measures for directed networks (digraphs) into AutoMap.....	125
Entropy.....	127
Reference .....	128
DYNetML .....	129
Changes in DyNetML outputs for XML files per Text.....	129
CASOS Email Parser (CEMAP).....	130
1. Instructions for the using the CASOS Email Parser (CEMAP) through AutoMap.	130
Introduction .....	130
Support.....	133
How to cite AutoMap .....	133
Further Information .....	133
Questions, Bugs, and Comments.....	133
Lesson 1 .....	135
From content analysis to semantic networks .....	135
I. Open AutoMap .....	135
On Empire: .....	135
II. Overview of the AutoMap Graphical User Interface (GUI).....	135
III. Load multiple text documents into AutoMap.....	136
Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextInputFiles .....	136



Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\	
TextInputFiles .....	136
IV. Carry out a simple content analysis.....	137
Note: The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts. ....	138
Create and refresh the Union Concept List: .....	138
Note: The number of unique concepts considers each concept only once. The number of total concepts also considers repetitions of concepts.....	140
V. Save your Union Concept List as a csv file.....	140
Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\	
UnionConceptList .....	141
VI. View your csv file in Microsoft Excel .....	141
Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\	
UnionConceptList .....	141
Note: You must have MS Excel installed on your computer in order to view your csv file this way. ....	142
VII. Save and view your results as a binary csv file.....	142
VIII. Carry out a simple Semantic Network Analysis.....	142
Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\	
TextOutputFiles.....	148
IX. Open ORA.....	149
X. Load your files into ORA .....	150
Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\	
TextOutputFiles.....	150
XI. Visualize your meta-matrices .....	150
XII. Run a Semantic Network Report .....	152
Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\	
SemanticNetworkReport.html.....	153
<i>SEMANTIC-NETWORK REPORT</i> .....	153
Symmetric difference .....	153
Saved output networks.....	154
Lesson 2 .....	154
Processing data: using the delete list .....	154
I. Open AutoMap .....	154
II. Overview of the AutoMap Graphical User Interface (GUI).....	154
III. Load multiple text documents into AutoMap.....	155
Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07. ....	155
Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07 .....	156
IV. Carry out a simple content analysis.....	156
Note: The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts. ....	157
Note: The number of unique concepts considers each concept only once. The number of total concepts also considers repetitions of concepts.....	159
V. The basics of pre-processing in AutoMap .....	159
VI. The basics of delete lists.....	159
VII. Open a Delete List.....	160

VIII. Delete concepts and examine the effects on the text and concept list .....	161
Note: If you wish to apply a Delete List and a Thesaurus, we recommend applying the Delete List first, followed by the Thesaurus. ....	161
IX. Un-apply a Delete List .....	162
X. Save an applied Delete List .....	162
Example: C:\Documents and Settings\carley\Desktop\Lesson 2\SampleDeleteList	163
XI. Save text(s) after application of Delete List.....	163
Example: C:\Documents and Settings\carley\Desktop\Lesson 2\TextsAfterDeletion .....	164
XII. Modify a Delete List .....	164
XIII. Create a delete list from scratch in AutoMap .....	164
Example: C:\Documents and Settings\carley\Desktop\Lesson 2\ SampleDeleteList.txt .....	164
Example: C:\Documents and Settings\carley\Desktop\Lesson 2\AppliedDeleteList .....	165
XIV. Open a chosen delete list with Microsoft Excel .....	165
XV. Modify a delete list in Microsoft Excel .....	165
Lesson 3 .....	166
Processing data: using a Generalization Thesaurus.....	166
I. Open AutoMap .....	166
II. Overview of the AutoMap Graphical User Interface (GUI) .....	166
III. Load multiple text documents into AutoMap.....	167
IV. Carry out a simple content analysis.....	167
Note: The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts. ....	168
Note: The number of unique concepts considers each concept only once. The number of total concepts also considers repetitions of concepts.....	169
V. The basics of pre-processing .....	169
VI. Open a Generalization Thesaurus.....	169
VII. Apply a thesaurus and examine its effect on the texts and concept list .....	170
Note: If you wish to apply both a Delete List and a Generalization Thesaurus, be sure to use first the Delete List and then the Thesaurus. ....	170
VIII. Un-apply a Generalization Thesaurus .....	171
IX. Modify the generalization thesaurus list .....	171
X. Create a generalization thesaurus list from scratch .....	171
XI. Save the Generalization Thesaurus .....	172
Example: C:\Documents and Settings\carley\Desktop\Lesson 3\AppliedGeneralizationThesaurus .....	172
XII. Open the Generalization Thesaurus in Microsoft Excel .....	173
XII. Modify the Generalization Thesaurus in Microsoft Excel.....	173
Lesson 4 .....	173
AutoMap Usage "Snapshots" .....	173
1. Snapshot: Split Input Text Files .....	173
2. Snapshot: Using Compare Maps .....	174
3. Snapshot: Merge DyNetML Files.....	174
4. Snapshot: Convert Network Data Formats .....	175

5. Snapshot: Edit Network Data .....	175
6. Snapshot: Visualize Semantic Networks .....	175
References .....	176
Where to learn more about Dynamic Network Analysis.....	178
Entropy references: .....	178

## **AutoMap: An Overview**

AutoMap is a software tool to analyze text using the method of Network Text Analysis. It performs a specific type of Network Text Analysis called Semantic Network Analysis. Semantic analysis extracts and analyzes links among words to model an authors “mental map” as a network of links. Additionally, Automap supports Content Analysis.

Coding in AutoMap is computer-assisted; the software applies a set of coding rules specified by the user in order to code the texts as networks of concepts. Coding texts as maps focuses the user on investigating meaning among texts by finding relationships among words and themes.

The coding rules in AutoMap involve text pre-processing and statement formation, which together form the coding scheme. Text pre-processing condenses data into concepts, which capture the features of the texts relevant to the user. Statement formation rules determine how to link concepts into statements.

Listed below are the steps a user would follow (in typical order) to use AutoMap:

1. Pre-Process texts.
2. Perform Semantic Network Analysis on texts.
3. Run MetaMatrix Text Analysis and Sub Matrix Text Analysis (Both techniques are sub-types of Map Analysis).
4. Compare Maps generated with AutoMap.
5. Compute network analytic measures per texts and words.

## **Network Text Analysis (NTA)**

NTA theory is based on the assumption that language and knowledge can be modeled as networks of words and relations. Network Text Analysis encodes links among words to construct a network of linkages. Specifically, Network Text Analysis analyzes the existence, frequencies, and covariance of terms and themes, thus subsuming classical Content Analysis.

## **Semantic Network Analysis**

In map analysis, a concept is a single idea, or ideational kernel, represented by one or more words. Concepts are equivalent to nodes in Social Network Analysis (SNA). The link between two concepts is referred to as a statement, which corresponds with an edge in SNA. The relation between two concepts can differ in strength, directionality, and type. The union of all statements per texts forms a semantic map. Maps are equivalent to networks.

## Social Network Analysis (SNA)

Social Network Analysis is a scientific area focused on the study of relations, often defined as social networks. In its basic form, a social network is a network where the nodes are people and the relations (also called links or ties) are a form of connection such as friendship. Social Network Analysis takes graph theoretic ideas and applies them to the social world. The term “social network” was first coined in 1954 by J. A. Barnes (see: *Class and Committees in a Norwegian Island Parish*). Social network analysis is also called network analysis, structural analysis, and the study of human relations. SNA is often referred to as the science of “connecting the dots.”

Today, the term Social Network Analysis (or SNA) is used to refer to the analysis of any network such that all the nodes are of one type (e.g., all people, or all roles, or all organizations), or at most two types (e.g., people and the groups they belong to). The metrics and tools in this area, since they are based on the mathematics of graph theory, are applicable regardless of the type of nodes in the network or the reason for the connections.

For most researchers, the nodes are actors. As such, a network can be a cell of terrorists, employees of global company or simply a group of friends. However, nodes are not limited to actors. A series of computers that interact with each other or a group of interconnected libraries can comprise a network also.

Where to find out more on SNA?

- *Scott, John, 2000, Social Networks, Sage (2nd edition)*
- *Wasserman, S. & K. Faust, 1994, Social Network Analysis: Methods and Applications.*

### Dynamic Network Analysis

Dynamic Network Analysis (DNA) is an emergent scientific field that brings together traditional social network analysis (SNA), link analysis (LA) and multi-agent systems (MAS). There are two aspects of this field. The first is the statistical analysis of DNA data. The second is the utilization of simulation to address issues of network dynamics. DNA networks vary from traditional social networks in that are larger dynamic multi-mode, multi-plex networks, and may contain varying levels of uncertainty.

DNA statistical tools are generally optimized for large-scale networks and admit the analysis of multiple networks simultaneously in which, there are multiple types of nodes (multi-node) and multiple types of links (multi-plex). In contrast, SNA statistical tools focus on single or at most two mode data and facilitate the analysis of only one type of link at a time.

DNA statistical tools tend to provide more measures to the user, because they have measures that use data drawn from multiple networks simultaneously. From a computer

simulation perspective, nodes in DNA are like atoms in quantum theory, nodes can be, though need not be, treated as probabilistic. Whereas nodes in a traditional SNA model are static, nodes in a DNA model have the ability to learn. Properties change over time; nodes can adapt: A company's employees can learn new skills and increase their value to the network; Or, kill one terrorist and three more are forced to improvise. Change propagates from one node to the next and so on. DNA adds the critical element of a network's evolution and considers the circumstances under which change is likely to occur.

## **Illustrative problems that people in the DNA area work on -**

1. Developing metrics and statistics to assess and identify change within and across networks.
2. Developing and validating simulations to study network change, evolution, adaptation, decay...
3. Developing and validating formal models of network generation and evolution.
4. Developing and testing theory of network change, evolution, adaptation, decay...
5. Developing techniques to visualize network change overall or at the node or group level.
6. Developing statistical techniques to see whether differences observed over time in networks are due to simply different samples from a distribution of links and nodes or changes over time in the underlying distribution of links and nodes.
7. Developing control processes for networks over time.
8. Developing algorithms to change distributions of links in networks over time.
9. Developing algorithms to track groups in networks over time.
10. Developing tools to extract or locate networks from various data sources such as texts.
11. Developing statistically valid measurements on networks over time.
12. Examining the robustness of network metrics under various types of missing data.
13. Empirical studies of multi-mode multi-link multi-time period networks.
14. Examining networks as probabilistic time-variant phenomena.
15. Forecasting change in existing networks Identifying trails through time given a sequence of networks. Identifying changes in node criticality given a sequence of networks anything else related to multi-mode multi-link multi-time period networks.

## **Getting Started**

- [Dynamic Network Analysis](#)
- [Social Network Analysis](#)

## Dynamic Network Analysis

Dynamic Network Analysis (DNA) is an emergent scientific field that brings together traditional social network analysis (SNA), link analysis (LA) and multi-agent systems (MAS). There are two aspects of this field. The first is the statistical analysis of DNA data. The second is the utilization of simulation to address issues of network dynamics. DNA networks vary from traditional social networks in that are larger dynamic multi-mode, multi-plex networks, and may contain varying levels of uncertainty.

DNA statistical tools are generally optimized for large-scale networks and admit the analysis of multiple networks simultaneously in which, there are multiple types of entities (multi-entities) and multiple types of links (multi-plex). In contrast, SNA statistical tools focus on single or at most two mode data and facilitate the analysis of only one type of link at a time.

DNA statistical tools tend to provide more measures to the user, because they have measures that use data drawn from multiple networks simultaneously. From a computer simulation perspective, entities in DNA are like atoms in quantum theory, entities can be, though need not be, and treated as probabilistic. Whereas entities in a traditional SNA model are static, entities in a DNA model have the ability to learn. Properties change over time; entities can adapt: A company's employees can learn new skills and increase their value to the network; Or, kill one terrorist and three more are forced to improvise. Change propagates from one entity to the next and so on. DNA adds the critical element of a network's evolution and considers the circumstances under which change is likely to occur.

### ***Where to learn to more:***

- Kathleen M. Carley, 2003, “ Dynamic Network Analysis” in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Ronald Breiger, Kathleen Carley, and Philippa Pattison, (Eds.) Committee on Human Factors, National Research Council, National Research Council. Pp. 133-145, Washington, DC.
- Kathleen M. Carley, 2002, “ Smart Agents and Organizations of the Future” The Handbook of New Media. Edited by Leah Lievrouw and Sonia Livingstone, Ch. 12, pp. 206-220, Thousand Oaks, CA, Sage.
- Kathleen M. Carley, Jana Diesner, Jeffrey Reminga, Maksim Tsvetovat, 2005-forthcoming, Toward an Interoperable Dynamic Network Analysis Toolkit, DSS Special Issue on Cyberinfrastructure for Homeland Security: Advances in Information Sharing, Data Mining, and Collaboration Systems.

## Social Network Analysis (SNA)

Social Network Analysis is a scientific area focused on the study of relations, often defined as social networks. In its basic form, a social network is a network where the entities are people and the relations (also called links or ties) are a form of connection

such as friendship. Social Network Analysis takes graph theoretic ideas and applies them to the social world. The term social network was first coined in 1954 by J. A. Barnes (see: *Class and Committees in a Norwegian Island Parish*). Social network analysis is also called network analysis, structural analysis, and the study of human relations. SNA is often referred to as the science of connecting the dots.

Today, the term Social Network Analysis (or SNA) is used to refer to the analysis of any network such that all the entities are of one type (e.g., all people, or all roles, or all organizations), or at most two types (e.g., people and the groups they belong to). The metrics and tools in this area, since they are based on the mathematics of graph theory, are applicable regardless of the type of nodes in the network or the reason for the connections.

For most researchers, the entities are actors. As such, a network can be a cell of terrorists, employees of global company or simply a group of friends. However, entities are not limited to actors. A series of computers that interact with each other or a group of interconnected libraries can comprise a network also.

### ***Where to find out more on SNA***

- *Scott, John*, 2000, *Social Networks*, Sage (2nd edition)
- *Wasserman, S. & K. Faust*, 1994, *Social Network Analysis: Methods and Applications*



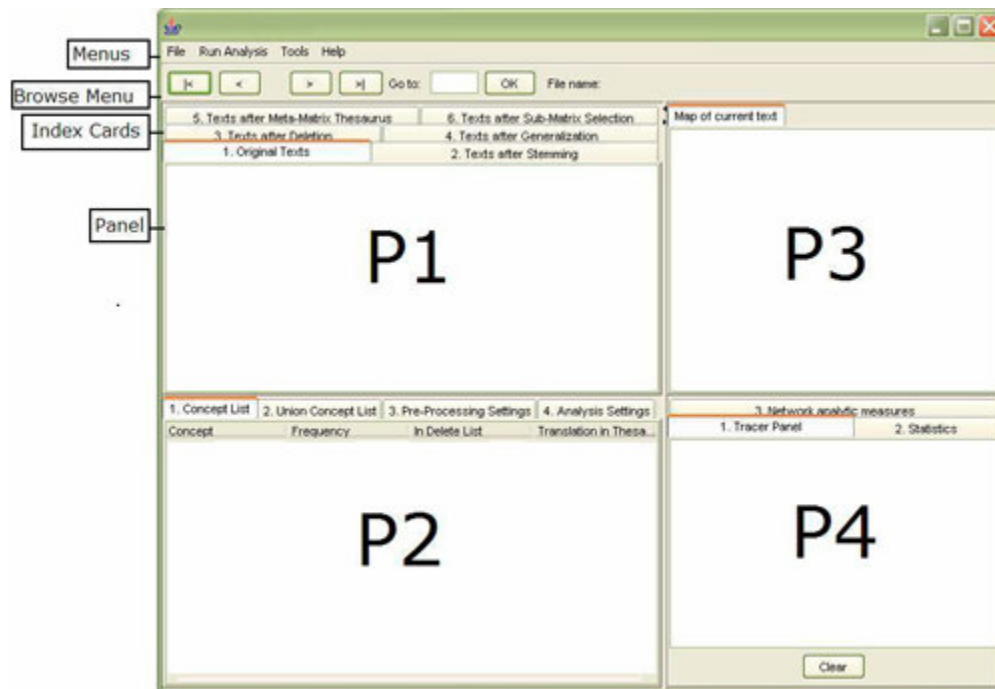
## Automap Graphical User Interface

AutoMap's graphical user interface (GUI) is divided into four primary quadrants (or panels): they will be referred to as P1 (top left), P2 (bottom left), P3 (top right) and P4 (bottom right).

The drop-down menu bar provides access to various analysis tools and utilities. The browse menu bar allows you to quickly navigate between loaded texts.

Index cards, or “tabs”, provide a tabular interface allowing you to navigate each panel respectively. The GUI reads any changes dynamically from the XML file. To do that, the user needs to refresh the tool.

The screen shot below highlights primary features of the Automap GUI and where to find them:



It is important to note that P2 can be edited. The other panels (P1, P3, P4) cannot be edited. Information displayed on P1 to P4 always relates to each other. The Text Browse Menu relates to all panels at the same time.

Window sizes do not have an upper threshold. AutoMap will automatically set window size to largest text size upon user's request. This is a new button on the Analysis Settings panel. This enables text set specific maximum window sizes, which also enhances efficiency.

The Action Tracer Panel in P4 will log preprocessing utilities applied to your text. This is a handy way to keep track of changes and actions relating to your text.

## ***Tool Tips***

In the various pre-processing panels, such as utilities, “tool tips”, provide more information on certain routines. Tool tips become visible when you slide the mouse over that particular tool.

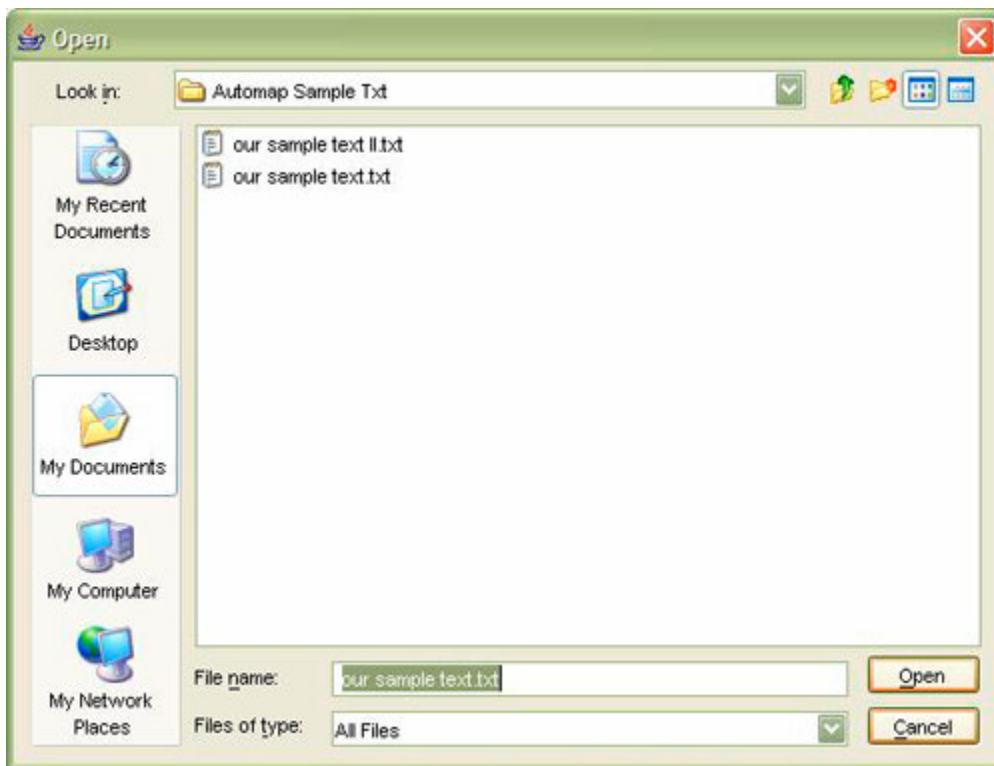
## **Load Input**

### ***1. To open a single text***

To open a single text file into Automap, proceed as follows from the Automap menu bar:

File Open > Open single file

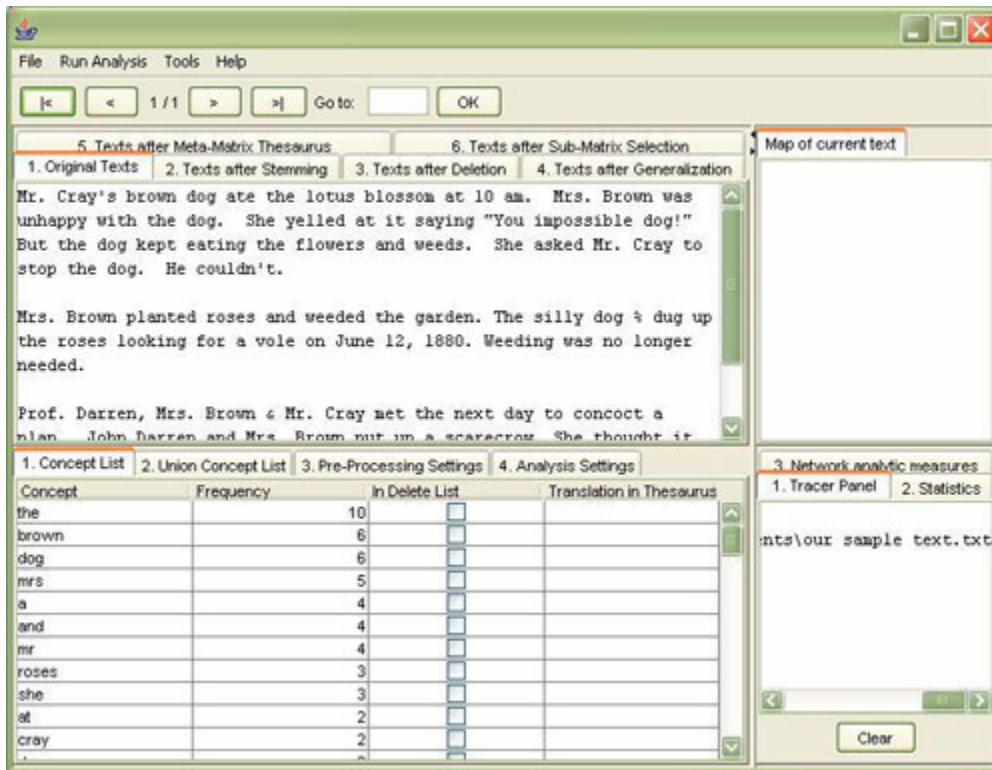
A file chooser will pop up (screen shot below).



Double click on the file that you wish to analyze > Select the Open button.

The text will be displayed in P1 on tab No. 1. Original Text.

The loading of .TXT files (caps) is now enabled.



## 2. To open a set of texts

Should you wish to analyze multiple texts at the same time, they must be stored in one folder. To do so, proceed as follows from the Automap menu bar:

File menu > Open multiple files

After you select “Open multiple files” a folder chooser will pop up. Again, be sure that correct folder is selected in the folder chooser. The black ellipses in the screen shot below highlight where you should be looking in the file chooser:

## Text Examples

This user's guide provides illustrative examples for all AutoMap functions. The sample texts below will be used throughout this guide.

**Tip!** It is suggested you follow along using the examples below. To do so, simply copy and paste the texts below into WordPad or notepad and save as a .txt file in the same folder.

## ***Our Text Example***

Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't.

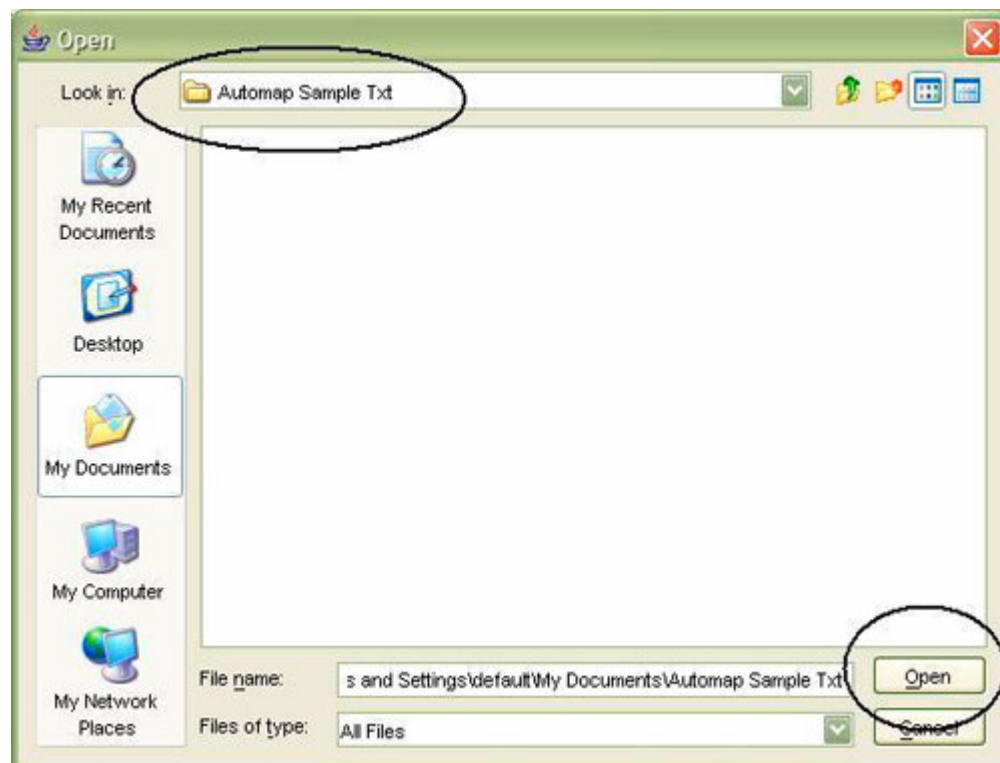
Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed.

Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Cray put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

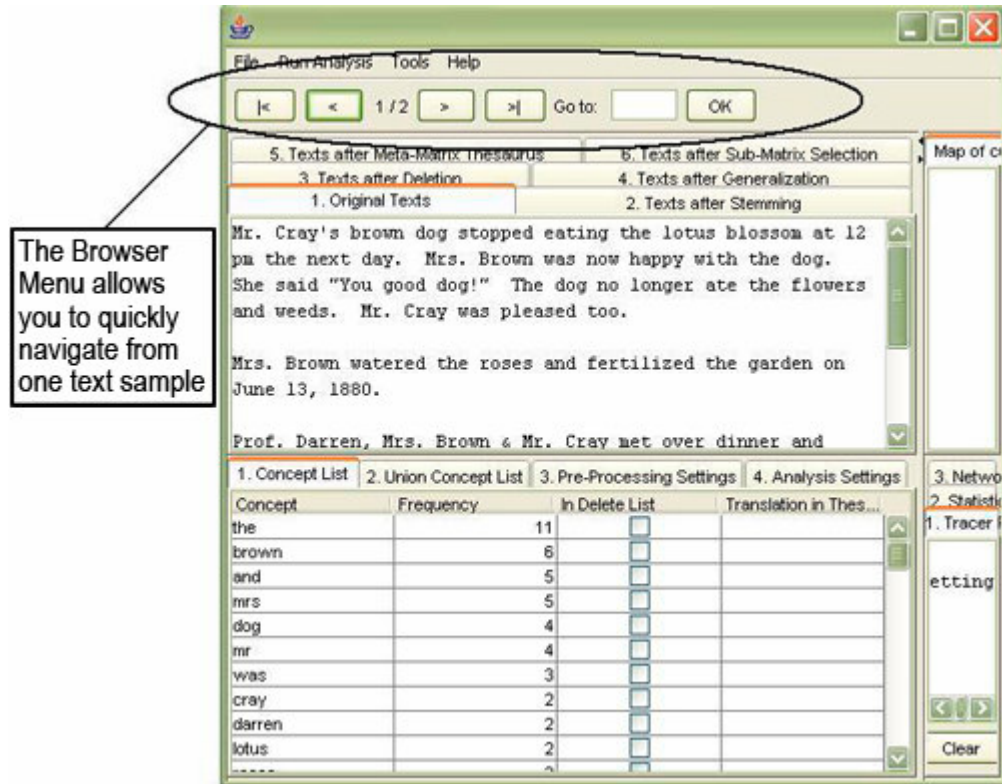
## ***Our Text Example II***

Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray was pleased too. Mrs. Brown watered the roses and fertilized the garden on June 13, 1880.

Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed how the plan had worked. John Darren and Mrs. Brown would take down the scarecrow the following week. She thought it was too scary for the dog. Mr. Cray painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.



Choose the folder that contains the texts you wish to analyze and single click on it. The folder will be highlighted. Do not double click on the folder. Select the Open button (see above screen shot). The first text will be displayed in panel P1, tab no. 1. Original Text. You can browse through the texts by using the Browse Menu.



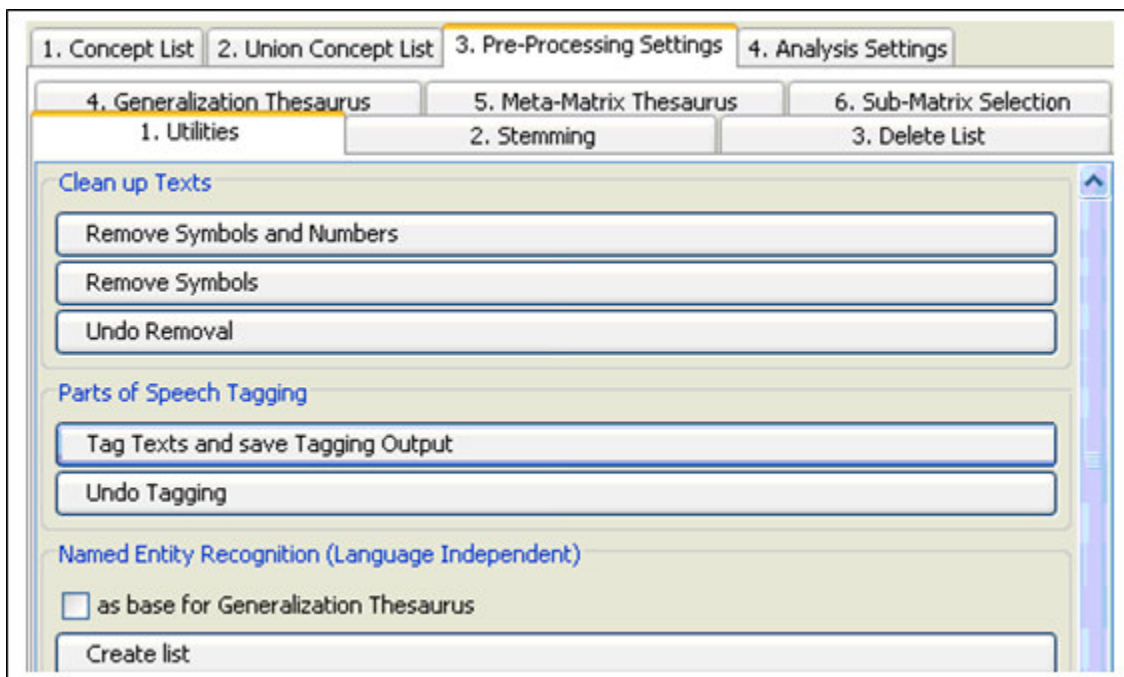
## Parts of Speech

The Parts of Speech feature does exactly that, when text is loaded into AutoMap this feature will show you what part of speech each word is.

To use the feature:

- In AutoMap menu bar, select File > Open single file. A folder chooser window will pop up.
- Navigate to the folder containing the texts you wish to analyze, and single click on it.
- Load your text file and it will appear in Panel 1 - Original Data.
- Next go to Panel 2 and select 3. Pre-Processing Settings.
- Once in the Pre-Processing Settings, select 1. Utilities.
- Scroll down the options and single click on Tag Texts and save Tagging Output, which is under the heading Parts of Speech Tagging.

**NOTE: Data has to be in txt format with .txt extension**



- The Tag Texts option that you have just selected will take your text document and define each part of text.
- Go back to Panel 1 and select 8. Texts after Parts-of-Speech Tagging. This option will show you your document with the different parts of speech.

Mr/MNP Cray's/MNP brown/JJ dog/MN ate/VBD the/DT lotus/MN blossom/MN at/IN 10/CD am/VBP Mrs/MNP Brown/MNP was/VBD unhappy/JJ with/IN the/DT dog/MN She/PRP yelled/VBD at/IN it/PRP saying/VBG "You/MN impossible/JJ dog!"/MN But/CC the/DT dog/MN kept/VBD eating/VBG the/DT flowers/MNS and/CC weeds/MNS She/PRP asked/VBD Mr/MNP Cray/MNP to/TO stop/VB the/DT dog/MN He/PRP couldn't/MD Mrs/MNP Brown/MNP planted/VBD roses/MNS and/CC weeded/VBN the/DT garden/MN The/DT silly/JJ dog/MN %/MN dug/VBN up/IN the/DT roses/MNS looking/VBG for/IN a/DT vole/MN on/IN June/MNP 12,/CD 1880/CD Weeding/MNP was/VBD no/RB longer/RB needed/VBN Prof/MNP Darren,/MNP Mrs/MNP Brown/MNP &/MN Mr/MNP Cray/MNP met/VBD the/DT next/JJ day/MN to/TO concoct/MN a/DT plan/MN John/MNP Darren/MNP and/CC Mrs/MNP Brown/MNP put/VBD up/RP a/DT scarecrow/MN She/PRP thought/VBD it/PRP would/MD scare/VB the/DT dog/MN Mr/MNP Cray/MNP put/VB up/RP a/DT fence/MN Problem/MNP solved/VBD Then/RB Mrs/MNP Brown/MNP planted/VBD lotus,/MN carnations,/MN daffodils,/MN and/CC

- To undo the Parts of Speech Tag, simply go back to Panel 2, 1. Utilities option Parts of Speech Tagging and select the Undo option.

## ***2. Parts of Speech Tagging***

This routine associates every word after the highest level of pre-processing applied so far with its Parts of Speech.

In order to tag your texts, go to the Utilities Panel, Parts of Speech Tagging field, and press the Tag Texts button.

To see the resulting tagging, go to the upper left panel, and Select 8. Texts after Parts of Speech tagging.

To undo the tagging, go to the Utilities Panel, Parts of Speech Tagging field, and press the Undo Tagging button.

This POS tagger was implemented based on a Hidden Markov Model. The learning data stems from the Penn Treebank 3 corpus. We are grateful to Alex Rudnicky from CMU for providing the training data to us.

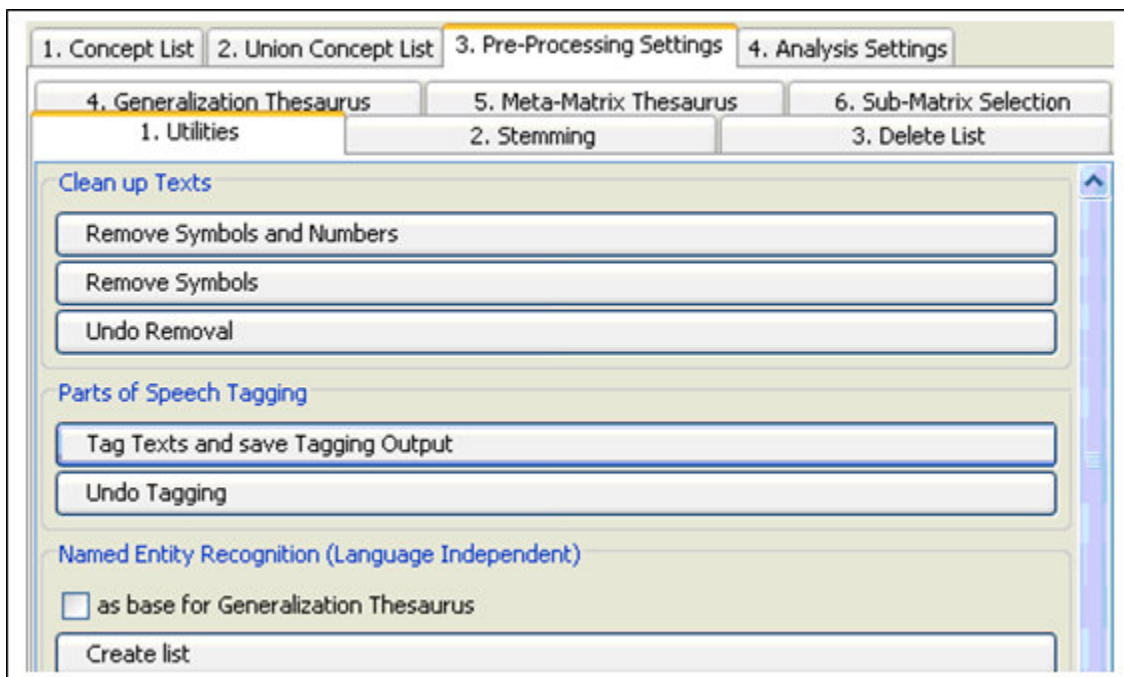
## Parts of Speech

The Parts of Speech feature does exactly that, when text is loaded into AutoMap this feature will show you what part of speech each word is.

To use the feature:

- In AutoMap menu bar, select File > Open single file. A folder chooser window will pop up.
- Navigate to the folder containing the texts you wish to analyze, and single click on it.
- Load your text file and it will appear in Panel 1 - Original Data.
- Next go to Panel 2 and select 3. Pre-Processing Settings.
- Once in the Pre-Processing Settings, select 1. Utilities.
- Scroll down the options and single click on Tag Texts and save Tagging Output, which is under the heading Parts of Speech Tagging.

**NOTE: Data has to be in txt format with .txt extension**



- The Tag Texts option that you have just selected will take your text document and define each part of text.
- Go back to Panel 1 and select 8. Texts after Parts-of-Speech Tagging. This option will show you your document with the different parts of speech.



```
Mr/MNP Cray's/MNP brown/JJ dog/MN ate/VBD the/DT lotus/MN
blossom/MN at/IN 10/CD am/VBP Mrs/MNP Brown/MNP was/VBD
unhappy/JJ with/IN the/DT dog/MN She/PRP yelled/VBD at/IN it/PRP
saying/VBG "You/MN impossible/JJ dog!"/MN But/CC the/DT dog/MN
kept/VBD eating/VBG the/DT flowers/MNS and/CC weeds/MNS She/PRP
asked/VBD Mr/MNP Cray/MNP to/TO stop/VB the/DT dog/MN He/PRP
couldn't/MD Mrs/MNP Brown/MNP planted/VBD roses/MNS and/CC
weeded/VBN the/DT garden/MN The/DT silly/JJ dog/MN %/MN dug/VBN
up/IN the/DT roses/MNS looking/VBG for/IN a/DT vole/MN on/IN
June/MNP 12,/CD 1880/CD Weeding/MNP was/VBD no/RB longer/RB
needed/VBN Prof/MNP Darren,/MNP Mrs/MNP Brown/MNP &/MN Mr/MNP
Cray/MNP met/VBD the/DT next/JJ day/MN to/TO concoct/MN a/DT
plan/MN John/MNP Darren/MNP and/CC Mrs/MNP Brown/MNP put/VBD
up/RP a/DT scarecrow/MN She/PRP thought/VBD it/PRP would/MD
scare/VB the/DT dog/MN Mr/MNP Cray/MNP put/VB up/RP a/DT
fence/MN Problem/MNP solved/VBD Then/RB Mrs/MNP Brown/MNP
planted/VBD lotus,/MN carnations,/MN daffodils,/MN and/CC
```

- To undo the Parts of Speech Tag, simply go back to Panel 2, 1. Utilities option Parts of Speech Tagging and select the Undo option.

## ***2. Parts of Speech Tagging***

This routine associates every word after the highest level of pre-processing applied so far with its Parts of Speech.

In order to tag your texts, go to the Utilities Panel, Parts of Speech Tagging field, and press the Tag Texts button.

To see the resulting tagging, go to the upper left panel, and Select 8. Texts after Parts of Speech tagging.

To undo the tagging, go to the Utilities Panel, Parts of Speech Tagging field, and press the Undo Tagging button.

This POS tagger was implemented based on a Hidden Markov Model. The learning data stems from the Penn Treebank 3 corpus. We are grateful to Alex Rudnicky from CMU for providing the training data to us.

## **Text Pre-Processing**

### ***1. Introduction to Text Pre-Processing in AutoMap***

Pre-processing reduces the data to terms relevant to you.

## **Tip! All pre-processing techniques in AutoMap are optional.**

There are some points to consider before you begin Pre-Processing:

1. Namely, Map Analysis can be run without any prior data pre-processing.
2. Meta Matrix Text Analysis and Sub Matrix Text Analysis **require** pre-processing.

Pre-processing is semi-automated and iterative and involves several key processes:

### ***Named-Entity Recognition***

Named-Entity Recognition is an Automap feature that allows you to retrieve proper names (e.g. names of people, organizations, places), numerals, and abbreviations from texts (Magnini, Negri, Prevete & Tanev, 2002). The AutoMap Named-Entity Recognition functionality detects:

- Single words that are capitalized.
  - Example: Copenhagen.
- Adjacent words that are capitalized.
  - Example: The New York City Police Department.
- A string of adjacent words that are capitalized, but can be intervened by one non-capitalized word. The first and the last word in this string are capitalized.
  - Example: Canadian Department of National Defense.

### ***Stemming***

Stemming detects inflections and derivations of concepts in order to convert each concept into the related morpheme (Jurafsky & Martin, p.83, 654). AutoMap offers 2 stemmers:

- The Porter Stemmer (Porter, 1980). This stemmer uses the Porter Stemming algorithm. Additionally, it converts irregular verbs into the verb's infinitive.
  - Example: The phrase "Hospitals switched to using emergency generators, will be stemmed to "Hospit switch to be us emerg gener."
- KSTEM or KROVETZ stemmer (Krovetz, 1995). An inflection and derivation-based stemmer. The KSTEM or KROVETZ stemmer can be customized.

### ***Collocation Identification***

A word's collocates are words appearing next to or near to it.

## **Tip! Collocations occurring with high frequency are powerful indicators of a pattern of meaning in a text.**

Collocations are helpful to construct thesauri in AutoMap. AutoMap can identify collocations of size 2 (Bigrams) as shown below:

- mobile phones

- shopping centre

## ***Deletion***

Deletion removes non-content bearing conjunctions and articles from texts (Carley, 1993). Non-content bearing concepts to be deleted from the texts are denoted in a Delete List. When applying a Delete List, AutoMap searches the text(s) for concepts specified in the Delete List and delete matches from the text(s). Example:

- Original input text: The New York City Police Department said a number of people were trapped in elevators for awhile.
- Entries in the Delete List: the, a, of, were, in, for, awhile.
- Text after deletion: New York City Police Department said number people trapped elevators.

## ***Thesauri***

A thesaurus associates concepts with more abstract concepts. When applying a thesaurus, AutoMap searches the text set for the text-level concepts denoted in the thesaurus and translates matches into the corresponding concept. The terminology of a thesaurus depends on the content and the subject of the data set (Burkart, 1997: 163; Zuell & Alexa, 2001: 313).

### ***Generalization Thesaurus.***

A generalization thesaurus typically is a two-columned collection that associates text-level concepts with higher-level concepts. The text-level concepts represent the content of a data set, and the higher-level concepts represent the text-level concepts in a generalized way (Burkart, 1997; Klein 1997: 256; Popping & Roberts 1997: 382).

- Example: Copenhagen will be associated with the higher-level concept City.
- Related type of Analysis: Map Analysis.

### ***Meta-Matrix Thesaurus***

A Meta-Matrix Thesaurus associates text-level concepts with meta-matrix categories. Since one concept might need to be translated into several meta-matrix categories, a meta-matrix thesaurus can consist of more than two columns. For example, the concept "commander" corresponds with the categories agent and knowledge.

- Example: City will be associated with and translated into Location.
- Related type of Analysis: Meta-Matrix Analysis.

For the meta-matrix thesaurus, column headers start with concept knowledge. The order AND naming of column headers of the meta-matrix thesaurus can be changed in the XML file.

## ***Sub-Matrix Selection.***

The Sub-Matrix Selection denotes which Meta-Matrix Categories should be retranslated into concepts used as input for the meta-matrix thesaurus.

- Example: Location will be selected and translated into Copenhagen, Oskarshamn and Ringhals, if those concepts were associated with the Meta-Matrix category "Location" in the Meta-Matrix Thesaurus.
- Related type of Analysis: Sub-Matrix Analysis.

## **Text Pre-Processing**

### ***2. Hierarchy of Pre-Processing Techniques***

If you apply a pre-processing technique of a lower order prior to a technique of higher order, the pre-processing will be maintained through all following procedures of higher order. You can un-apply each technique after applying it, if needed.

### **Tip! All pre-processing techniques are optional.**

If you wish to apply multiple pre-processing techniques, do this in the following order:

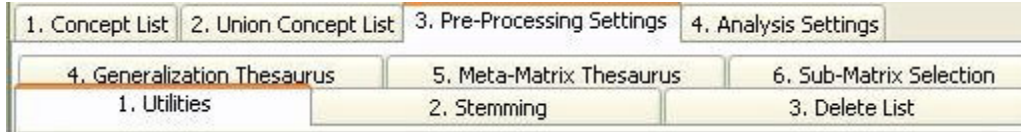
1. Named-Entity Recognition: This is an utility that does not impact the data. Can be used before any type of analysis is run. Can be used before or after Stemming.
2. Collocation/ bigram Identification: This is an utility that does not impact the data. Can be used before any type of analysis is run.
3. Stemming: Can be used before any type of analysis is run. Can be used before or after Named-Entity Recognition.
4. Deletion: Can be used before any type of analysis is run.
5. Thesauri:
  1. Generalization Thesaurus  
Can be applied before Semantic Network Analysis is run. Can be applied before Meta-Matrix Thesaurus is applied.
  2. Meta-Matrix Thesaurus  
Has to be applied if Meta-Matrix Analysis should be run.
  3. Sub-Matrix Selection  
Can only be performed if Meta-Matrix Thesaurus was applied. Has to be applied if Sub-Matrix Analysis should be run.

The numbering of the index card tabs on P1 and P2 reflect this hierarchy in order to make the sequence of the pre-processing steps more intuitive.

Numbering of index card tabs on P1:

6. Texts after Meta-Matrix Thesaurus	7. Texts after Sub-Matrix Selection
4. Texts after Deletion	5. Texts after Generalization
1. Original Texts	2. Texts after Symbol Removal
	3. Texts after Stemming

Numbering of index card tabs on P2:

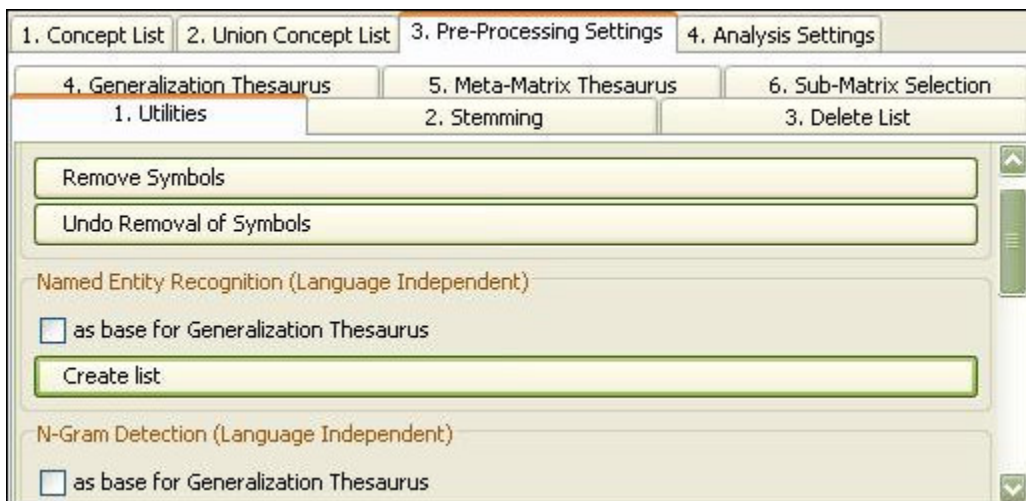


## Text Pre-Processing

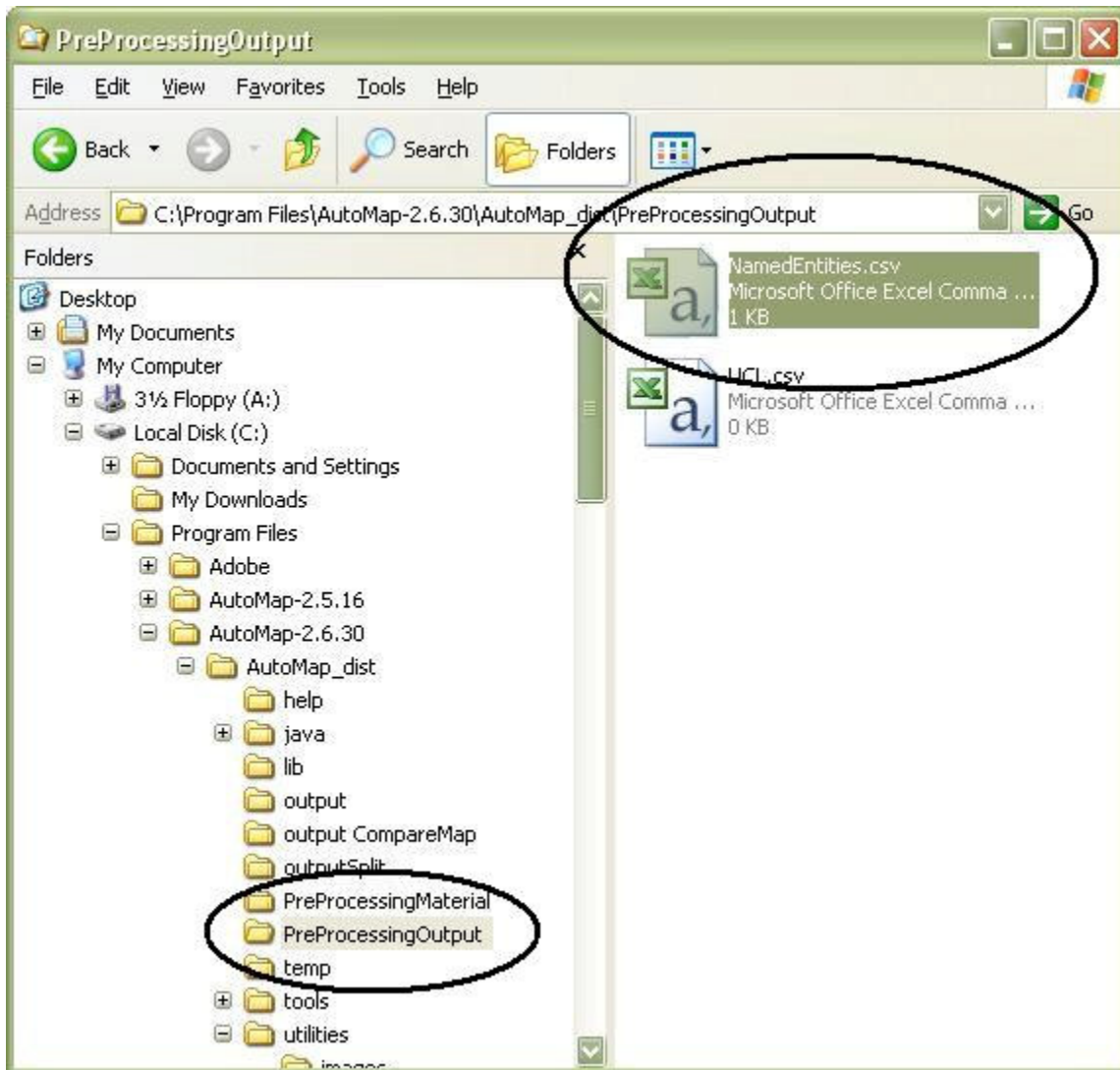
### 3.1 Named-Entity Recognition

To create a list of all Name-Entities that are contained in the data set opened, go to Utilities (tab no. 3) in P2 and click the Create and save Named Entities List button in the Named-Entities Field. The resulting list will be automatically saved under NamedEntities.csv in the root directory of AutoMap.

The Named-Entity Recognition interface:



The black ellipses in the screen shot below highlight where to find the NamedEntities.csv file in your root directory.



### 3.1.1 Example for Named-Entity Recognition

Resulting NamedEntities.csv file after the following Small Predefined Delete List was applied with rhetorical adjacency:

	John Darren and Mrs. Brown
a	Mr. Cray
an	He couldn't. Mrs. Brown
some	Prof. Darren Mrs. Brown & Mr. Cray
many	13-Jun
this	But
that	Mr. Cray's
these	Mr. Craye
those	Problem solved. Then Mrs. Brown
the	Then Mrs. Brown
all	12-Jun
one	Weeding
every	She
	She asked Mr. Cray

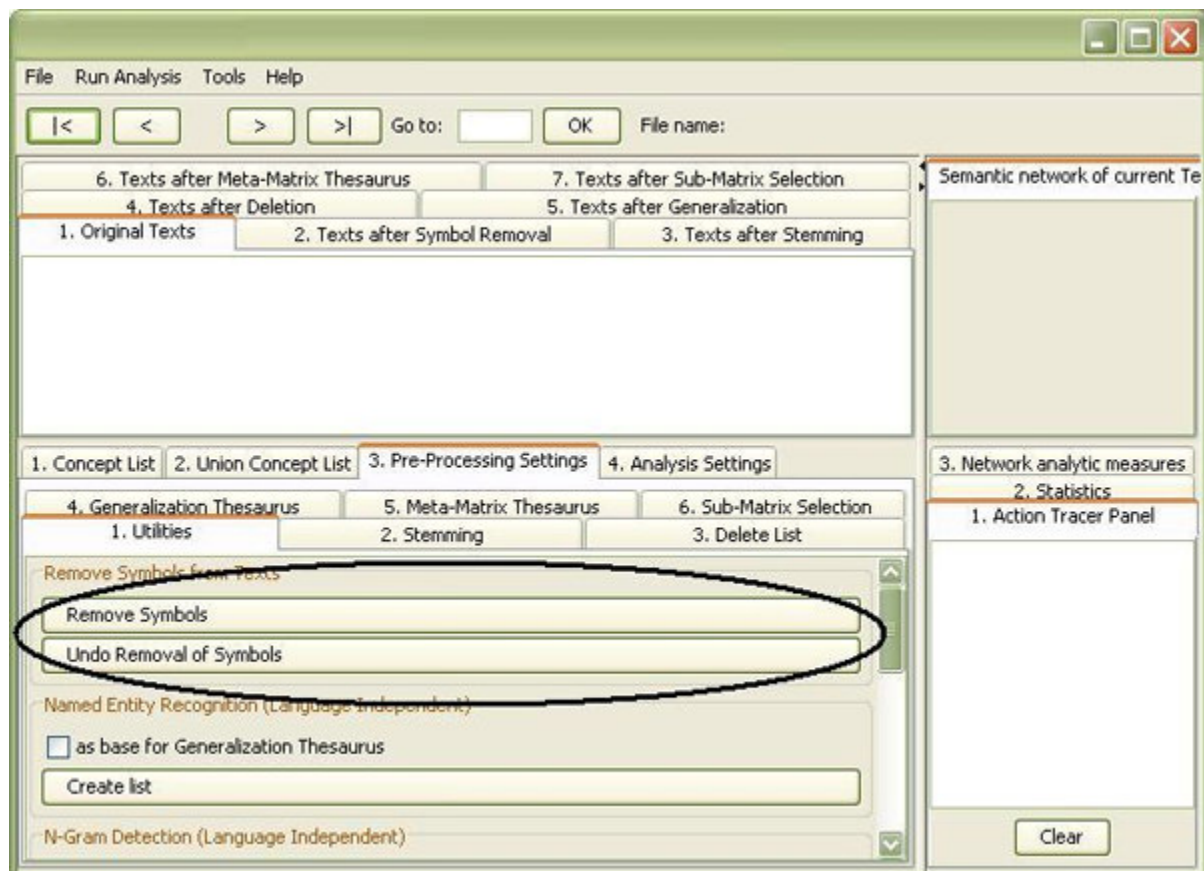
Mrs. Brown

Redundant concepts can be converted to one word by stemming. Concepts not relevant to the user can be eliminated by deletion.

### 3.2 Symbol Removal

In the lower left panel (P2) you will find an option under tab no. 3 Pre-Processing Utilities labeled on tab no. 1 as Symbol Removal. This routine removes or strips off all characters that are neither a letter nor a number. It maintains sentence marks. It converts question marks and exclamation marks into sentence marks. This helps replace the delchar option on the delete list in a more user-friendly fashion. The overall purpose of this routine is to do a very thorough cleaning of the data in a fully automated, easy to use fashion. This routine can be unapplied by using the un-apply button, which is located close the apply button.

### Symbol Removal



### **3.21 N-gram Identification: Bigrams**

To create a list of all bigrams that are contained in the data set opened, go to the Utilities, tab no. 3, in P2 and select the Create Bigram (Correlation) List button in the N-gram Detection. The resulting list will be automatically saved under CorrelationList.csv in the root directory of AutoMap.

## **Text PreProcessing**

### **4. Stemming**

To stem a text (or text set), go to the Stemming (tab no. 4) in P2.

#### **Porter Stemmer:**

To apply the Porter Stemmer select the Apply button next to Porter Stemmer and stemming of irregular verbs for English. The stemmed text(s) will be displayed on the tab no. 2 Stemmed Text in P1.

To unstem the texts, go to tab no. 4. Stemming in P2 and select the Un-Apply button. Tab no. 2. Stemmed Text in P1 will be cleared.

#### **Krovetz Stemmer:**

For the Krovetz stemmer, several customization options are offered:

1. Decide whether capitalized words should be stemmed or not.  
Use radio buttons in the interface to make your selection. By default, capitalized words are stemmed.
2. Define words to be modified by the stemmer.  
These words are collected in a protection list, named `selfdefined_protected_concepts.txt`, stored in the AutoMap root directory under `utilities\KStem`. To avoid stemming certain words put them in this list, one word per line, without any line delimiter.
3. Define specific stems for certain words.  
These words are collected in a list of pairwise associations, named `selfdefined_pairs.txt`, which are stored in the AutoMap root directory under `utilities\KStem`. To stem a certain word into a pre-defined term, put the pair (first word / pre-defined stem) in the list, one pair per line, without any line delimiter. The `selfdefined_pairs.txt` list that comes along with AutoMap contains already such pairs, which handle the correct stemming or irregular verbs in English.

To apply the Krovetz stemmer:

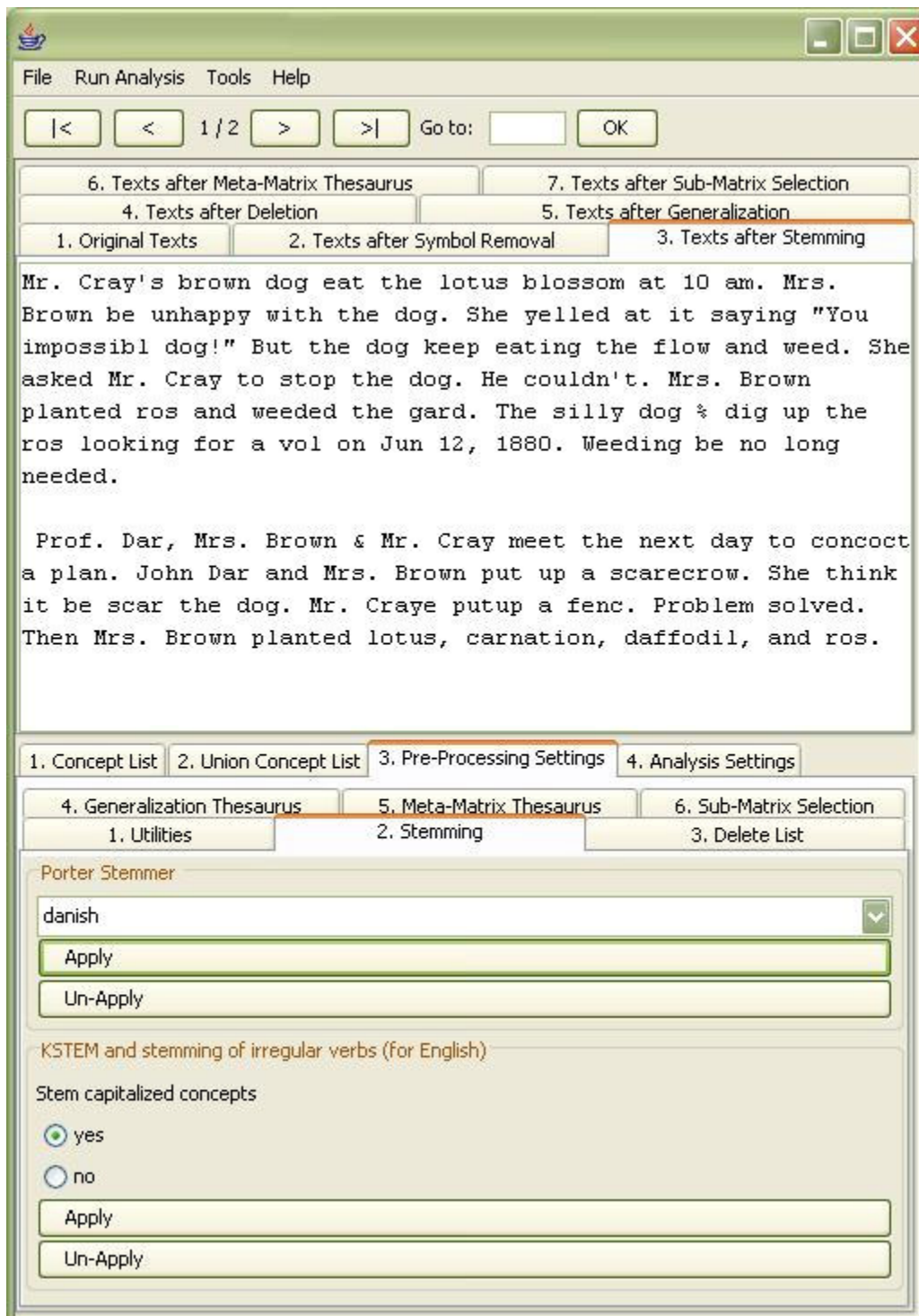
Select the Apply button next to K-stem. The stemmed text(s) will be displayed on tab no. 2. Stemmed Text index card in P1.



To unstem the texts, go to the tab no. 4 Stemming in P2 and select the Un-Apply button. The tab no. 2 Stemmed Text in P1 will be cleared.

#### ***4.1 Example for Stemming (Porter)***

Stemmed text in P1 and interface of Stemming index card on P2:



## Text PreProcessing

### 5. Deletion

The Delete List is not case sensitive.

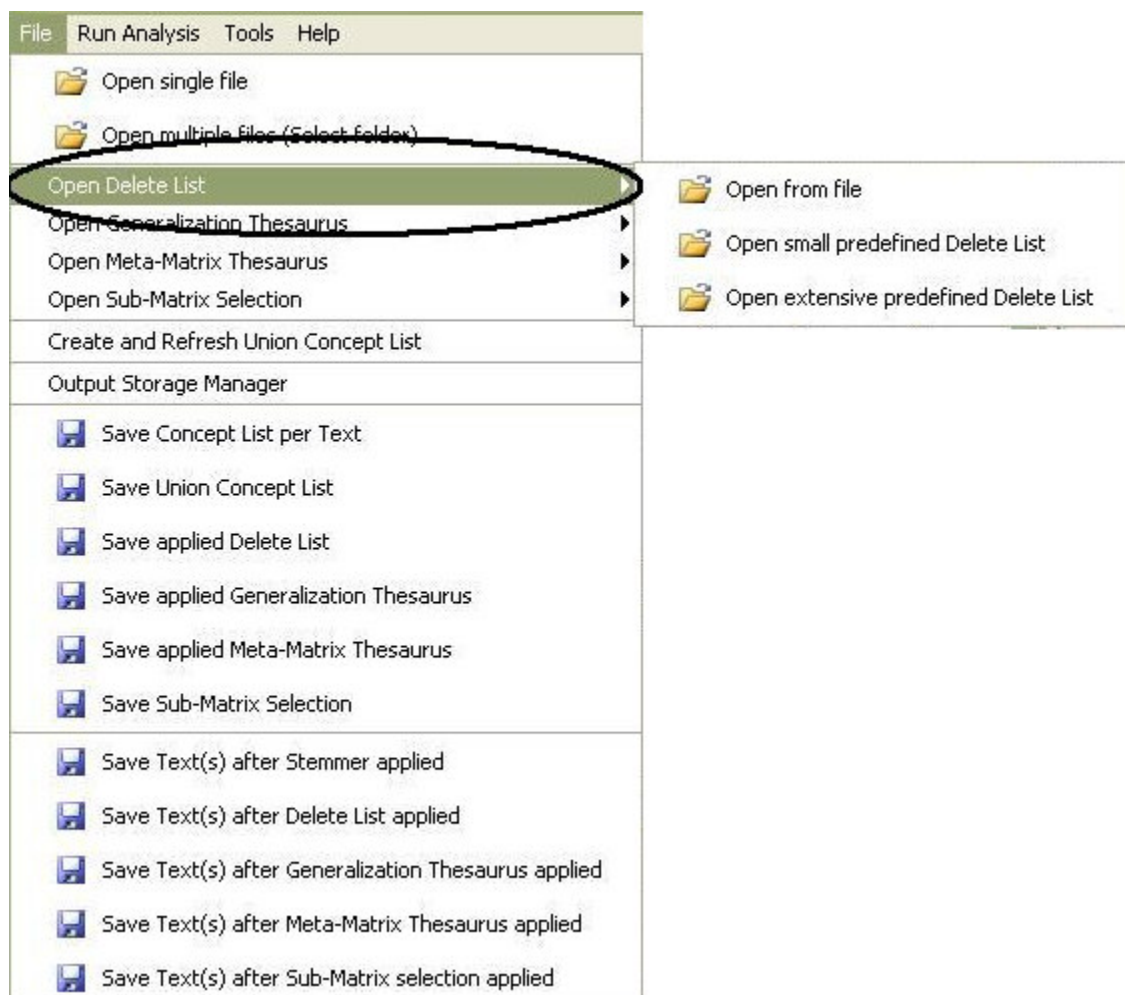
You can use the predefined Delete Lists that AutoMap offers or create your own Delete List. All lists can be edited.

## 5.1 Open a Delete List

Click the File menu, select Open Delete List and choose one of the following options:

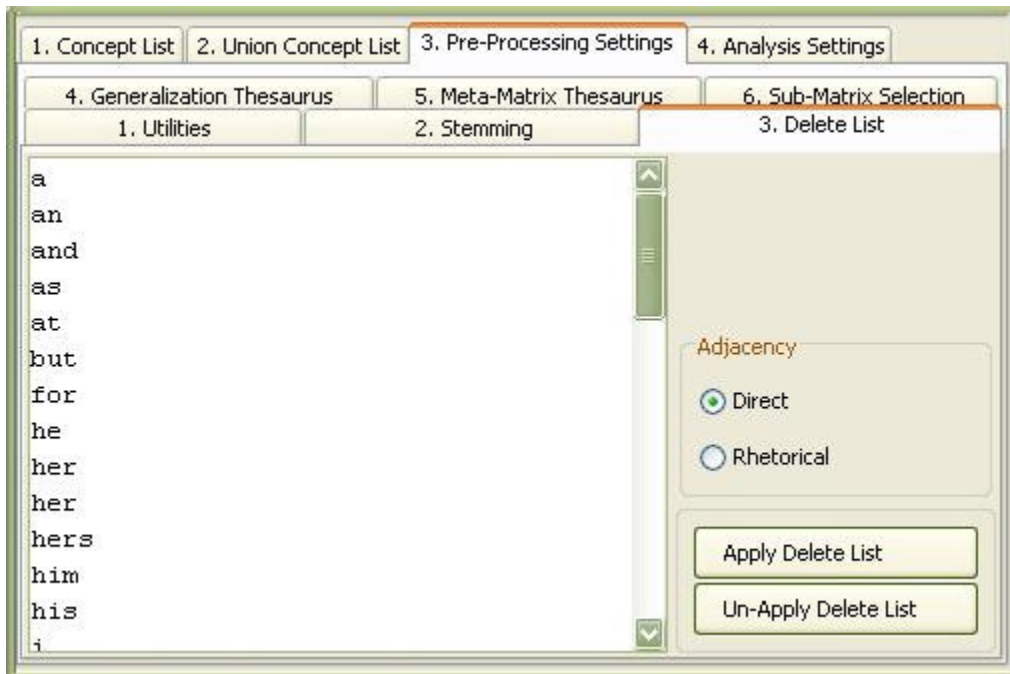
- Open from file: A file chooser will appear. Select a delete list and hit the Open button.
- Open small predefined Delete List: AutoMap's predefined small delete list will be opened.
- Open extensive predefined Delete List. AutoMap's extensive small delete list will be opened.

The black ellipses in the screen shot below shows where to access the Delete List utility:



The Delete List will be displayed in P2, tab no. 5. Delete List index card:

Interface of the Delete List index card:



The Delete List can be edited. (see section 5.3 Edit a Delete List)

### ***5.1.1 Small predefined Delete List***

The Delete List is compiled of words that occur most frequently in English: a, an, and, some, many, this, that, these, those, the, all, one, every.

The Small Delete List can be edited. (see section 5.3 Edit a Delete List)

### ***5.1.2 Extensive predefined Delete List***

An Extensive Delete List is based on words occurring most frequently in English: a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself. As the name indicates, the Extensive Delete list contains more words than the Small Predefined Delete List.

The Extensive Delete List can be edited. (see section 5.3 Edit a Delete List)

## ***5.2 Create a Delete List***

There are two ways to create a Delete List:

1. Within AutoMap:  
Go to the Delete List index card. The general structure of a Delete List is one single concept per line. Add concepts by typing one concept per line. Hit enter

after entering a concept. Avoid empty lines. See the example for more information.

- 2.
3. Outside of AutoMap:  
Use a text editor to create a Delete List. Please consider these instructions to create a Delete List:
  1. The general structure of a Delete List is one single Concept per line.
  2. Avoid empty lines.
  3. The Delete List is NOT case sensitive.
  4. Save the List.
  5. Open the Delete List in AutoMap.
  6. You can edit the Delete List in AutoMap if you wish.

### ***5.3. Edit a Delete List***

On the Delete List index card you can:

- Add concepts: Type one concept per line. Hit enter after entering a concept.
- Modify concepts: Go to the line and retype concept.
- Drop concepts: Mark the concept and hit the delete key.

### ***5.4 Apply a Delete List***

If you wish to apply a Delete List and a Thesaurus we recommend first applying a Delete List and then a Thesaurus. Next, follow these steps:

1. Before applying a delete list, an adjacency option can be chosen on the Delete List index card. Adjacency can be either direct (default) or rhetorical. If the user does not change the adjacency option, AutoMap uses direct adjacency for deletion and analysis.
2. To delete the concepts specified in the Delete List from all texts loaded click the Apply Delete List button on the Delete Concepts Index card.
3. See the pre-processed texts in P1, Delete List (tab no. 3).

When applying a Delete List AutoMap does three things:

1. Search the text(s) for concepts specified in the Delete List.
2. Delete matches from the text(s).
3. Display the resulting text(s) in P1, Delete List (tab no. 3).  
If direct adjacency was chosen, concepts specified in the delete list are simply deleted from texts and concepts left and right to deleted concepts will appear adjacent to each other in terms of visualization and statement formation.

If rhetorical adjacency was chosen placeholders (xxx) are inserted where a concept was deleted. The placeholders retain original distances of maintained concepts for purposes of visualization and analysis.

To apply multiple delete lists load the first one in, apply it, then load in the next, apply it, and so on.

### **5.4.1 Direct Adjacency**

If direct adjacency is chosen, concepts in the text that match concepts specified in the delete list will be deleted from texts. As a result concepts left and right of a deleted concept move together and will be treated as directly adjacent to each other for visualization and analysis.

To apply direct adjacency check the radio button in the Delete List index card. Then apply the delete list. If the user does not change the adjacency option, AutoMap uses direct adjacency for deletion and analysis.

### **5.4.2 Rhetorical adjacency**

If rhetorical adjacency is chosen placeholders "xxx" are inserted where a concept was deleted. The placeholders retain the original distances of the maintained concepts visually for analysis.

To apply direct adjacency check the button on the Delete List tab. Then apply the delete list.

If the user does not change the adjacency option, AutoMap uses direct adjacency for deletion and analysis.

## **5.5 Un-Apply a Delete List**

To un-apply a Delete List that was applied to the data, in P2 go to the Delete List (tab no. 5) index card and select the Un-Apply button. The tab no. 3 Delete List index card on P1 will be cleared.

## **5.6 Save an applied Delete List**

To save a Delete List that you have applied to the data, click the File menu, select Save Delete List as. A file chooser will pop up.

## **5.7 Save text(s) after application of Delete List**

To save the text(s) after the application of the Delete List, click the File menu, select Save Text(s) after Delete List applied. All texts are automatically saved in a folder called "preprocessed" in the root directory of AutoMap. The filename will be "after\_DL\_NameOfYourText.txt".

## **5.8 Examples for the application of a Delete List**

<b>Input text</b>	<b>Tool used</b>	<b>Setting</b>	<b>Resulting text</b>
Mr. Cray's brown dog ate the lotus blossom at	AutoMap's extensive	Direct Adjacency	Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was

	(default)	unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.
	Delete List:	
	<i>a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself</i>	
10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.	Rhetorical Adjacency	Mr. Cray's brown dog ate xxx lotus blossom xxx 10 am. Mrs. Brown was unhappy with xxx dog. xxx yelled xxx xxx saying "You impossible dog!" xxx xxx dog kept eating xxx flowers xxx weeds. xxx asked Mr. Cray xxx stop xxx dog. xxx couldn't. Mrs. Brown planted roses xxx weeded xxx garden. xxx silly dog % dug up xxx roses looking xxx xxx vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met xxx next day xxx concoct xxx plan. John Darren xxx Mrs. Brown put up xxx scarecrow. xxx thought xxx xxx scare xxx dog. Mr. Craye put up xxx fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, xxx roses.
	The following concepts were added to the Delete List. Direct Adjacency (default)	
	<i>from, in, what, was, with</i>	Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown unhappy dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.
	Rhetorical	Mr. Cray's brown dog ate xxx lotus

Adjacency blossom xxx 10 am. Mrs. Brown xxx  
unhappy xxx xxx dog. xxx yelled xxx  
xxx saying "You impossible dog!"  
xxx xxx dog kept eating xxx flowers  
xxx weeds. xxx asked Mr. Cray xxx  
stop xxx dog. xxx couldn't. Mrs.  
Brown planted roses xxx weeded xxx  
garden. xxx silly dog % dug up xxx  
roses looking xxx xxx vole on June  
12, 1880. Weeding xxx no longer  
needed. Prof. Darren, Mrs. Brown &  
Mr. Cray met xxx next day xxx  
concoct xxx plan. John Darren xxx  
Mrs. Brown put up xxx scarecrow.  
xxx thought xxx xxx scare xxx dog.  
Mr. Cray put up xxx fence. Problem  
solved. Then Mrs. Brown planted  
lotus, carnations, daffodils, xxx  
roses.

## **Text PreProcessing**

### ***6. Generalization Thesaurus***

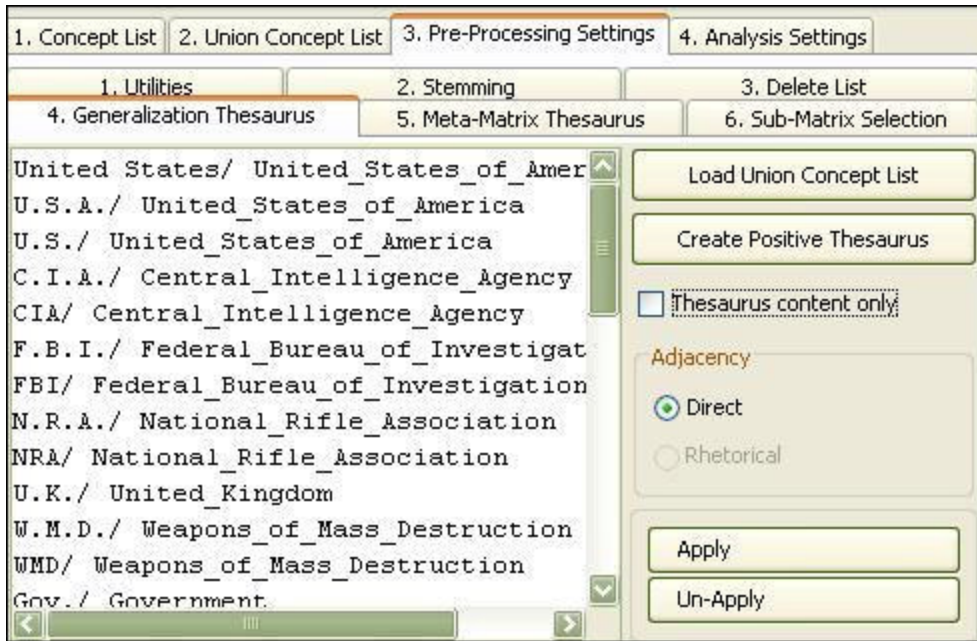
The Generalization Thesaurus is NOT case sensitive.

#### ***6.1. Open a Generalization Thesaurus***

Click the File menu, select Open Generalization Thesaurus. A file chooser will pop up. Double click the thesaurus you wish to wish to open or single click the thesaurus and then hit the Open button. The thesaurus will be displayed on P2 6. Generalization Thesaurus index card.

Generalization Thesaurus Interface index card:





## 6.2. Create a Generalization Thesaurus

There are two ways to create a Thesaurus:

### 1. Within AutoMap:

Go to P2, (tab no. 6) Generalization Thesaurus (see also the interface of the Generalization Thesaurus index card for an example).

Use the Text Area on this Index card.

Build and edit a thesaurus.

AutoMap supports users in building a generalization thesaurus by loading the union of concepts from the highest level of pre-processing applied into the Generalization Thesaurus field. This is found on the Generalization Thesaurus index card and can be used upon demand.

Follow these steps to load the union concept list into the Generalization Thesaurus field:

1. Create or refresh the Union Concept List.
2. Hit the Load Union Concept List button on the Generalization Thesaurus index card.

This concept list loaded into AutoMap can be refined by applying Named-Entity Recognition and Deletion prior to Generalization.

## ***Here is an example for multi-step pre-processing:***

To further illustrate multi-step pre-processing techniques, copy the text passages below, then save as a TXT file as "Our Text I.txt" and "Our Text II.txt" respectively. Load these examples into Automap to follow along.

Input texts:

Our Text I.txt

Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

Our Text II.txt:

Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray was pleased too. Mrs. Brown watered the roses and fertilized the garden on June 13, 1880. Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed how the plan had worked. John Darren and Mrs. Brown would take down the scarecrow the following week. She thought it was too scary for the dog. Mr. Craye painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.

<b>Pre-processing technique applied</b>	<b>Entries</b>	<b>Result</b>
<b>1. Create Named-Entity List</b>		John Darren and Mrs. Brown Mr. Cray He couldn't. Mrs. Brown Prof. Darren Mrs. Brown & Mr. Cray 13-Jun But Mr. Cray's Mr. Craye Problem solved. Then Mrs. Brown Then Mrs. Brown 12-Jun Weeding She She asked Mr. Cray Mrs. Brown
<b>2. Create Named-Entity</b>	Mr. Cray's/Mr. Craye Mr.	

**List and use it to build a generalization thesaurus.**

**3. Add further words that belong together to the generalization thesaurus.**

Craye/Mr. Cray's

Prof.  
Darren/Prof\_Darren

**4. Apply generalization thesaurus (no thesaurus content only).**

mr. craye brown dog ate the lotus blossom at 10 am. mrs. brown was unhappy with the dog. she yelled at it saying "you impossible dog!" but the dog kept eating the flowers and weeds. she asked mr. cray to stop the dog. he couldn't. mrs. brown planted roses and weeded the garden. the silly dog % dug up the roses looking for a vole on june 12, 1880. weeding was no longer needed.

mrs. brown & mr. cray, mrs. brown & mr. cray met the next day to concoct a plan. john darren and mrs. brown put up a scarecrow. she thought it would scare the dog. mr. cray's put up a fence. problem solved. brown planted lotus, carnations, daffodils, and roses.

Mr. Cray's brown dog ate xxx lotus blossom at 10 am. Mrs. Brown was unhappy with xxx dog. She yelled at it saying "You impossible dog!" But xxx dog kept eating xxx flowers and weeds. She asked Mr. Cray to stop xxx dog. He couldn't. Mrs. Brown planted roses and weeded xxx garden. xxx silly dog % dug up xxx roses looking for xxx vole on June 12, 1880. Weeding was no longer needed.

**3. Deletion (rhetorical adjacency).**

an  
some  
many  
this  
that  
these  
those  
the  
all  
one  
every

Prof. Darren, Mrs. Brown & Mr. Cray met xxx next day to concoct xxx plan. John Darren and Mrs. Brown put up xxx scarecrow. She thought it would scare xxx dog. Mr. Craye put up xxx fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

**4. Create Union Concept List.**

(Coincides with Union Concept List loaded into AutoMap, see cell below)

**5. Load Union Concept List into AutoMap.**

1880  
asked  
blossom

brown  
but  
concoct  
craye  
daffodils  
dinner  
discussed  
dog  
dog!"  
down  
dug  
fence  
for  
his  
how  
it  
looking  
lotus  
on  
over  
plan  
pm  
problem  
said  
saying  
scary  
solved  
stopped  
take  
up  
watered  
weeded  
weeding  
weeds  
with  
worked  
yelled

Outside of AutoMap: Use a text editor to create a Thesaurus.

Build and edit a thesaurus.

Save the Thesaurus.

Open the Thesaurus in AutoMap.

You can edit the Thesaurus in AutoMap if you wish.

## **6.3 Edit a Generalization Thesaurus**

You can add, change or drop the lines of a thesaurus on P2, (tab no. 6) Generalization Thesaurus.

The general structure of a Thesaurus follows the five points below (see also the interface of the Generalization Thesaurus index card for an example):

1. Every line contains Concept / Key Concept or in other words Old Word/ New Word.
2. A Concept can be one or more words.
3. A Key Concept is one word.
4. Be sure to separate the words by a slash.
5. The Thesaurus is NOT case sensitive.

## **6.4 Apply a Generalization Thesaurus**

If you wish to apply a Delete List and a Generalization Thesaurus please be sure to use the Delete List first and then the Thesaurus. Then go through the following process:

1. Decide if you want to use the Thesaurus content only option or not. If you do not select the Thesaurus content only option this setting will not be applied.
2. If you select the Thesaurus content only option you can choose an adjacency option.
3. Adjacency can be either direct (default) or rhetorical.
4. To apply your Generalization Thesaurus with the settings you have specified click the Apply Thesaurus button on the Generalization Thesaurus Index card. AutoMap uses the entries in the Thesaurus to search the text(s) for concepts. If a match is found it will be translated into a key concept. Again, the Thesaurus is NOT case sensitive.
5. See the pre-processed texts on the P1, (tab no. 4) Generalization Thesaurus. If the Thesaurus content only option and Direct Adjacency were chosen only key concepts would be displayed and considered for analysis. If the Thesaurus content only option and Rhetorical Adjacency were chosen key concepts and their original distances, which are symbolized by place holders (xxx), are displayed and considered for analysis.

### **6.4.1 Thesaurus content only**

If the Thesaurus content only option is chosen AutoMap performs the following steps:

1. Search the text(s) for concepts specified in the thesaurus.
2. Translate matches into key concepts.
3. Maintain only key concepts in the pre-processed texts. The rest of the input text is dropped and will not be considered for further pre-processing or analysis. The original distances of the key concepts will not be maintained. However, punctuation marks like the end of sentences and paragraphs are maintained and considered for analysis.
4. As a result, all key concepts in the resulting text appear directly adjacent to each other.

To select the Thesaurus content only option check the Thesaurus content only item on P2, (tab no. 6.) Generalization Thesaurus index card. You can now choose to either use direct or rhetorical adjacency for the application of the Generalization Thesaurus. Then apply the Generalization Thesaurus. To switch from not using the Thesaurus content only option uncheck the Thesaurus content only item on P2, (tab no. 6.) Generalization Thesaurus and apply the Generalization Thesaurus again.

If the Thesaurus content only option is NOT chosen AutoMap performs the following steps:

1. Search the text(s) for concepts specified in the thesaurus.
2. Translate matches into key concepts.
3. Keep the rest of the text as it is. This means, all other concepts in the text that did not match concepts specified in the thesaurus will not be affected in any way. Original distances of both unaffected concepts and key concepts will be maintained. This rule does not apply if a concept consisting of more than one word was translated into a key concept.

The Thesaurus content only item on P2, (tab no. 6) Generalization Thesaurus index card by default is not checked because AutoMap does not apply the Thesaurus content only option. To switch to using the Thesaurus content only option check the Thesaurus content only item on P2, (tab no. 6) Generalization Thesaurus and then apply the Generalization Thesaurus again.

#### **6.4.1.1 Direct Adjacency**

Direct adjacency means that original distances of concepts that represent the key concepts will neither be visualized nor considered for analysis.

To choose the direct adjacency click the Direct button in the Adjacency field on P2, (tab no. 6) Generalization Thesaurus index card. Then apply the Generalization Thesaurus. If the user does not change the adjacency option, AutoMap uses direct adjacency for generalization and analysis.

#### **6.4.1.2 Rhetorical Adjacency**

Rhetorical adjacency means that the original distance of key concepts will be considered for the analysis. Original distances of concepts that represent the key concepts will be visually symbolized by placeholders (xxx) and considered for analysis. Rhetorical adjacency can only be applied if the Thesaurus content only option was not chosen.

To choose the rhetorical adjacency click the Rhetorical button in the Adjacency field on P2, Generalization Thesaurus (tab no. 6). Then apply the Generalization Thesaurus.

If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.

## 6.5 Un-Apply a Generalization Thesaurus

To un-apply a Generalization Thesaurus that was applied to the data, go to P2, Generalization Thesaurus (tab no. 6) and hit the Un-Apply button. The Generalization Thesaurus (Tab no. 4) on P1 will be cleared.

## 6.6 Save an applied Generalization Thesaurus

To save a Generalization Thesaurus that you have applied to the data, click the File menu, select Save Generalization Thesaurus As (a file chooser will pop up).

## 6.7 Save text(s) after application of Generalization Thesaurus

To save the text(s) after the application of the Generalization Thesaurus, click the File menu, select Save Text(s) after Generalization Thesaurus applied. All texts are automatically saved in a folder called "preprocessed" in the root directory of AutoMap. The filename will be after\_general\_thes\_NameOfYourText.txt".

## 6.8 Example for the building and applying a Generalization Thesaurus

Applying a thesaurus to text that was not pre-processed

Input text	Tool used	Setting	Resulting text
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the	Mr Craye/Mr_Cray Mrs Browns/Mrs_Brown John Darren/Prof_John_Darren Prof Darren/Prof_John_Darren yelled/yelling ate/eating	Not Selected - Thesaurus content only (default)	mr. cray's brown dog eating the lotus blossom at 10 am. mrs. brown was unhappy with the dog. she yelling at it saying "you impossible dog!" but the dog kept eating the flowers and weeds. she asked mr. cray to stop the dog. he couldn't. mrs. brown planted roses and weeded the garden. the silly dog % dug up the roses looking for a vole on june 12, 1880. weeding was no longer needed. prof. darren, mrs. brown & mr. cray met the next day to concoct a plan. prof_john_darren and mrs. brown put up a

flowers  
and  
weeds.  
She asked  
Mr. Cray  
to stop the  
dog. He  
couldn't.  
Mrs.  
Brown  
planted  
roses and  
weeded  
the  
garden.  
The silly  
dog % dug  
up the  
roses  
looking  
for a vole  
on June  
12, 1880.  
Weeding  
was no  
longer  
needed.

Prof.  
Darren,  
Mrs.  
Brown &  
Mr. Cray  
met the  
next day  
to concoct  
a plan.  
John  
Darren  
and Mrs.  
Brown put  
up a  
scarecrow.  
She  
thought it  
would

daffodils, and roses.

Thesaurus  
content only,  
Direct Adjacency  
(default if  
Thesaurus  
content only is  
chosen)

eating... yelling..... ,....  
prof\_john\_darren.....

Selected -  
Thesaurus  
content only,  
Rhetorical  
Adjacency

xxx. xxx xxx xxx eating xxx  
xxx xxx xxx xxx xxx. xxx. xxx  
xxx xxx xxx xxx xxx. xxx  
yelling xxx xxx xxx xxx  
xxx xxx xxx xxx xxx xxx  
xxx xxx xxx. xxx xxx xxx. xxx  
xxx xxx xxx xxx. xxx xxx. xxx.  
xxx xxx xxx xxx xxx xxx xxx.  
xxx xxx xxx xxx xxx xxx xxx  
xxx xxx xxx xxx xxx xxx, xxx.  
xxx xxx xxx xxx xxx.

xxx. xxx, xxx. xxx xxx. xxx  
xxx xxx xxx xxx xxx xxx xxx.  
prof\_john\_darren xxx xxx. xxx  
xxx xxx xxx. xxx xxx xxx xxx  
xxx xxx xxx. xxx. xxx xxx xxx.  
xxx xxx. xxx xxx. xxx xxx xxx,  
xxx, xxx, xxx xxx.



scare the  
dog. Mr.  
Craye put  
up a fence.  
Problem  
solved.  
Then Mrs.  
Brown  
planted  
lotus,  
carnations,  
daffodils,  
and roses.

***Applying a thesaurus to text that was pre-processed with a Delete List, direct adjacency:***

Input text	Tool used	Setting	Resulting text
Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12,	Thesaurus (same as above)	Not selected - Thesaurus content only (default)	mr. cray's brown dog eating lotus blossom 10 am. mrs. brown was unhappy with dog. yelling saying "you impossible dog!" dog kept eating flowers weeds. asked mr. cray stop dog. couldn't. mrs. brown planted roses weeded garden. silly dog % dug up roses looking vole on june 12, 1880. weeding was no longer needed. prof. darren, mrs. brown & mr. cray met next day concoct plan. prof_john_darren mrs. brown put up scarecrow. thought scare dog. mr. craye put up fence. problem solved. then mrs. brown planted lotus, carnations, daffodils, roses.
		Selected - Thesaurus content only, Direct Adjacency	eating... yelling.....,,
		(default if Thesaurus content only is chosen)	.,... prof_john_darren.....,,
		Thesaurus content only, Rhetorical Adjacency	xxx. xxx xxx xxx eating xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx. yelling xxx xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx. xxx xxx xxx. xxx. xxx. xxx xxx xxx

1880.  
Weeding  
was no  
longer  
needed.

Prof.  
Darren, Mrs.  
Brown &  
Mr. Cray  
met next day  
concoct  
plan. John  
Darren Mrs.  
Brown put  
up  
scarecrow.  
thought  
scare dog.  
Mr. Craye  
put up fence.  
Problem  
solved. Then  
Mrs. Brown  
planted  
lotus,  
carnations,  
daffodils,  
roses.

xxx xxx. xxx xxx xxx xxx xxx xxx  
xxx xxx xxx xxx, xxx. xxx xxx  
xxx xxx xxx.

xxx. xxx, xxx. xxx xxx. xxx xxx  
xxx xxx xxx xxx.  
prof\_john\_darren xxx. xxx xxx  
xxx xxx. xxx xxx xxx. xxx. xxx  
xxx xxx. xxx xxx. xxx xxx. xxx  
xxx xxx, xxx, xxx, xxx.

### ***Applying a thesaurus to text that was pre-processed with a Delete List, rhetorical adjacency:***

Input text	Tool used	Setting	Resulting text
Text after application of Customized extensive delete list, rhetorical adjacency:	Thesaurus (same as above)	Not Selected - Thesaurus content only (default)	mr. cray's brown dog eating xxx lotus blossom xxx 10 am. mrs. brown was unhappy with xxx dog. xxx yelling xxx xxx saying "you impossible dog!" xxx xxx dog kept eating xxx flowers xxx weeds. xxx asked mr. cray xxx stop xxx dog. xxx couldn't. mrs. brown planted roses xxx weeded xxx garden. xxx silly dog % dug up xxx roses looking xxx xxx vole on june 12, 1880. weeding was no longer
Mr. Cray's brown dog ate xxx lotus			

blossom xxx  
 10 am. Mrs.  
 Brown was  
 unhappy  
 with xxx  
 dog. xxx  
 yelled xxx  
 xxx saying  
 "You  
 impossible  
 dog!" xxx  
 xxx dog kept  
 eating xxx  
 flowers xxx  
 weeds. xxx  
 asked Mr.  
 Cray xxx  
 stop xxx  
 dog. xxx  
 couldn't.  
 Mrs. Brown  
 planted  
 roses xxx  
 weeded xxx  
 garden. xxx  
 silly dog %  
 dug up xxx  
 roses  
 looking xxx  
 xxx vole on  
 June 12,  
 1880.  
 Weeding  
 was no  
 longer  
 needed.

Prof.  
 Darren, Mrs.  
 Brown &  
 Mr. Cray  
 met xxx next  
 day xxx  
 concoct xxx  
 plan. John  
 Darren xxx

needed.

prof. darren, mrs. brown & mr.  
 cray met xxx next day xxx concoct  
 xxx plan. prof\_john\_darren xxx  
 mrs. brown put up xxx scarecrow.  
 xxx thought xxx xxx scare xxx  
 dog. mr. craye put up xxx fence.  
 problem solved. then mrs. brown  
 planted lotus, carnations, daffodils,  
 xxx roses.

Thesaurus content  
 only, Direct  
 Adjacency

(default if Thesaurus  
 content only is  
 chosen)

. eating... yelling.....,

.,... prof\_john\_darren.....,

Thesaurus content  
 only, Rhetorical  
 Adjacency

xxx. xxx xxx xxx eating xxx xxx  
 xxx xxx xxx xxx. xxx. xxx xxx  
 xxx xxx xxx xxx. xxx yelling xxx  
 xxx xxx xxx xxx xxx xxx xxx  
 xxx xxx xxx xxx xxx xxx. xxx xxx  
 xxx. xxx xxx xxx xxx xxx. xxx  
 xxx. xxx. xxx xxx xxx xxx xxx  
 xxx xxx. xxx xxx xxx xxx xxx xxx  
 xxx xxx xxx xxx xxx xxx xxx xxx,  
 xxx. xxx xxx xxx xxx xxx.

xxx. xxx, xxx. xxx xxx. xxx xxx  
 xxx xxx xxx xxx xxx xxx xxx.  
 prof\_john\_darren xxx xxx. xxx  
 xxx xxx xxx xxx. xxx xxx xxx xxx  
 xxx xxx xxx. xxx. xxx xxx xxx  
 xxx. xxx xxx. xxx xxx. xxx xxx  
 xxx, xxx, xxx, xxx xxx.

Mrs. Brown  
put up xxx  
scarecrow.  
xxx thought  
xxx xxx  
scare xxx  
dog. Mr.  
Craye put up  
xxx fence.  
Problem  
solved. Then  
Mrs. Brown  
planted  
lotus,  
carnations,  
daffodils,  
xxx roses..

## Text PreProcessing

### ***7. Meta-Matrix Thesaurus***

A Meta-Matrix Thesaurus has to be applied if Meta-Matrix Analysis should be performed.

A Meta-Matrix Thesaurus associates concepts with meta-matrix categories:

- Agent
- Knowledge
- Resource
- Task/Event
- Organization
- Location
- Action
- Role
- Attribute
- Any user-defined category (as many as the user defines)

When applying a Meta-Matrix Thesaurus, AutoMap searches the text(s) for the entries specified in the Meta-Matrix Thesaurus and translates matches into related Meta-Matrix categories. If you also want to apply a Delete List or / and a Generalization Thesaurus you will need to apply these pre-processing tools before the Meta-Matrix Thesaurus. The Meta-Matrix Thesaurus is NOT case sensitive.

You might also see the meta matrix model as implemented in AutoMap to better understand the meta-matrix.

## **7.1 Open a Meta-Matrix Thesaurus**

Click the File menu, select Open Meta-Matrix Thesaurus and choose Open from highest level of pre-processing.

The union of concepts after the highest level of pre-processing applied so far will be displayed in alphabetical order on P2, Meta-Matrix Thesaurus (tab no. 7) index card.

### **Note: The Meta-Matrix Thesaurus can be edited.**

If you have a pre-defined Meta-Matrix Thesaurus available that matches (some of) the concepts contained in the loaded Meta-Matrix Thesaurus you can open this file.

To do so, click the File menu, select Open Meta-Matrix Thesaurus and choose Open from file.

If a concept contained in the pre-defined file matches a concept in the currently opened Meta-Matrix Thesaurus the meta-matrix categories assigned to this concept in the pre-defined file will be automatically assigned to the concept in the currently opened Meta-Matrix Thesaurus.

### **Note: The pre-assigned Meta-Matrix Thesaurus can be edited.**

## **7.2 Edit a Meta-Matrix Thesaurus**

To each concept that appears in the Concept column of the Meta-Matrix Thesaurus you can assign special categories:

- One or multiple meta-matrix categories to a concept by checking these categories. To unselect an assignment, uncheck the meta-matrix categories.
- Self-defined meta-matrix categories (e.g. location for Copenhagen). To do this, enter the category you wish to define in the last column (User-defined) of the Meta-Matrix Thesaurus. You can define as many categories as you want to. You can assign a concept to multiple self-defined categories. In this case, you will need a single row for each assignment. Hit enter when you have finished a self-defined row. To unselect an assignment, uncheck the meta-matrix categories.
- No meta-matrix category to a concept. In this case, this concept will not be considered for meta-matrix pre-processing and Meta-Matrix Analysis.

## **7.3 Build a Meta-Matrix Thesaurus**

You can build a Meta-Matrix Thesaurus outside of AutoMap by using a text editor. Please consider these instructions:

1. Every line contains (Key) Concept / Meta-Matrix category / Meta-Matrix category/.... You can assign a (key) concept to as many Meta-Matrix

categories as you want to. If a concept that appears in the Meta-Matrix Thesaurus written by you does not appear in the Concepts column of the Meta-Matrix Thesaurus this will not result in a conflict. AutoMap automatically searches for concepts contained in the Concepts column of the Meta-Matrix Thesaurus and when it finds a concept that also appears in your manually built Thesaurus it assigns the categories that you assigned to this concept to the concept on the index card.

2. Avoid empty lines.
3. Make sure to separate the words by a slash.
4. The Meta-Matrix Thesaurus is NOT case sensitive.
5. Save the Meta-Matrix Thesaurus.
6. Open the Meta-Matrix Thesaurus in AutoMap.
7. You can edit the Meta-Matrix Thesaurus in AutoMap if you wish.

## **7.4 Apply a Meta-Matrix Thesaurus**

Meta-matrix pre-processing is a higher level of pre-processing than the application of a Delete List and a Generalization Thesaurus. Thus, if you also want to apply a Delete List or / and a Generalization Thesaurus you will need to apply these pre-processing tools before the Meta-Matrix Thesaurus.

Follow this process:

1. Decide if you want to use the Thesaurus content only option or not.  
If you do not select the Thesaurus content only option this setting will not be applied.
2. If you select the Thesaurus content only option you can choose an adjacency option.
3. Adjacency can be either direct (default) or rhetorical.  
If you do not change the adjacency option, no adjacency option will be applied.
4. To apply your Meta-Matrix Thesaurus with the settings you have specified click the Apply Thesaurus button on the Meta-Matrix Thesaurus Index card. AutoMap uses the entries in the Thesaurus to search the text(s) for concepts. If a match is found it will be translated into a Meta-Matrix category. The Thesaurus is case insensitive.
5. See the pre-processed texts on the P1, Meta-Matrix Thesaurus (tab no. 5). Meta-Matrix categories are confined by tags (e.g., <agent>) . If multiple Meta-Matrix categories were assigned to a concept these categories appear tagged (e.g., if police was assigned to agent and resource, than the translated text would look like this: <agent, resource>).

### **7.4.1 Thesaurus content only**

If the Thesaurus content only option is chosen AutoMap does the following:

1. Search the text(s) for concepts specified in the Concept column of the Meta-Matrix Thesaurus.
2. Translate matches into related meta-matrix categories.

3. Maintain only meta-matrix categories in the pre-processed texts. The rest of the text is dropped and will not be considered for further pre-processing or analysis. Punctuation marks like the end of sentences and paragraphs are maintained and considered for analysis.

To select the Thesaurus content only option check the Thesaurus content only item on P2, Meta-Matrix Thesaurus (tab no. 7). You can now choose to either direct (default) or rhetorical adjacency. Then apply the Meta-Matrix Thesaurus. In order to switch from using the Thesaurus content only option uncheck the Thesaurus content only item and apply the Thesaurus again.

If the Thesaurus content only option is NOT chosen AutoMap does the following:

1. Search the text(s) for concepts specified in the thesaurus.
2. Translate matches into meta-matrix categories.
3. Keep the rest of the text as it is. This means, all other concepts in the text that did not match concepts specified in the thesaurus will not be affected at all. Original distances of both unaffected concepts and meta-matrix categories will be maintained. This rule does not apply if a concept consisting of more than one word is translated into a meta-matrix category.

AutoMap by default does not select the Thesaurus content only option. Therefore, the Thesaurus content only item on P2, Meta-Matrix Thesaurus (tab no. 7) by default is not checked. Just apply the Meta-Matrix Thesaurus. In order to switch to using the Thesaurus content only option check the Thesaurus content only item on P2, Meta-Matrix Thesaurus (tab no. 7) index card and then apply the Thesaurus again.

#### **7.4.1.1 Direct Adjacency**

Direct adjacency means that original distances of concepts that represent the key concepts will neither be visualized nor considered for analysis.

To choose the direct adjacency click the Direct button in the Adjacency field on P2, Meta-Matrix Thesaurus index card (tab no. 7). Then apply the Meta-Matrix Thesaurus. If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis. As a result, only meta-matrix categories are displayed on P1, Meta-Matrix Thesaurus (tab no. 5) will be considered for analysis. All meta-matrix tags in the resulting text appear directly adjacent to each other.

#### **7.4.1.2 Rhetorical Adjacency**

Rhetorical adjacency can only be applied if the Thesaurus content only option was not chosen.

To choose the rhetorical adjacency click the Rhetorical button in the Adjacency field on P2, (tab no. 7) Meta-Matrix Thesaurus. Then apply the Meta-Matrix Thesaurus. If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.

As a result, the meta-matrix tags and the rest of the text are displayed on P1, (tab no. 5) Meta-Matrix Thesaurus and will be considered for analysis. Original distances of meta-matrix tags that represent the key concepts will be visually symbolized by placeholders (xxx) and considered for analysis.

### ***7.5 Un-Apply a Meta-Matrix Thesaurus***

To un-apply a Meta-Matrix Thesaurus that was applied to the data, go to P2, (tab. no. 7) Meta-Matrix Thesaurus and hit the Un-Apply button. The tab no. 5 Meta-Matrix Thesaurus on P1 will be cleared.

### ***7.6 Save an applied Meta-Matrix Thesaurus***

If you wish to save a Meta-Matrix Thesaurus you first need to apply it. To save the Thesaurus, click the File menu, select Save Meta-Matrix Thesaurus as. A file chooser will pop up.

### ***7.7 Save text(s) after application of Meta-Matrix Thesaurus***

To save the text(s) after the application of the Meta-Matrix Thesaurus, click the File menu, select Save Text(s) after Meta-Matrix Thesaurus applied. All texts are automatically saved in a folder called "preprocessed" in the root directory of AutoMap. The filename will be after\_MMCATheS\_NameOfYourText.txt".

### ***7.8 Example for editing and applying a Meta-Matrix Thesaurus***

An Extract from the Our Text I.txt was used as input:

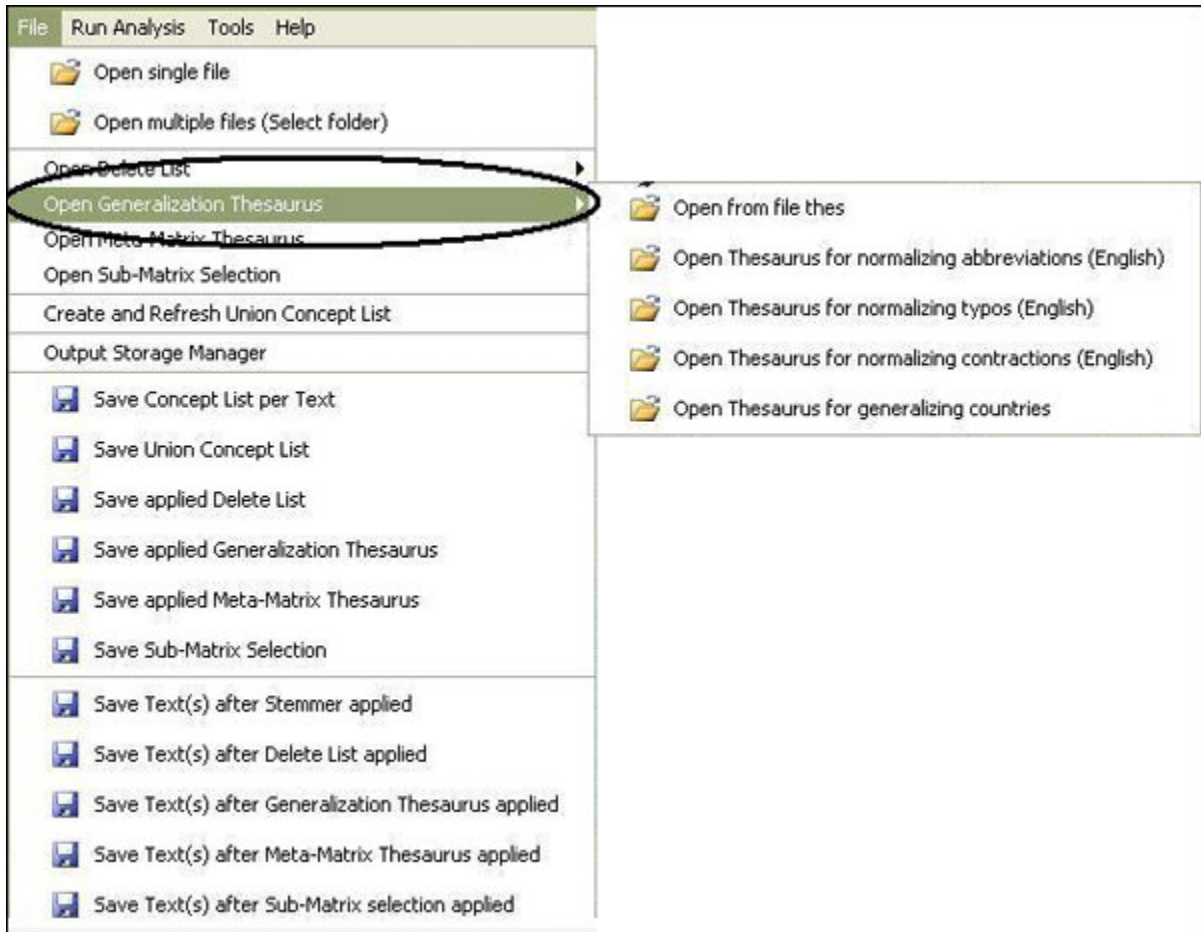
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

The customized extensive Delete List was applied to this text. The resulting text looks like this:

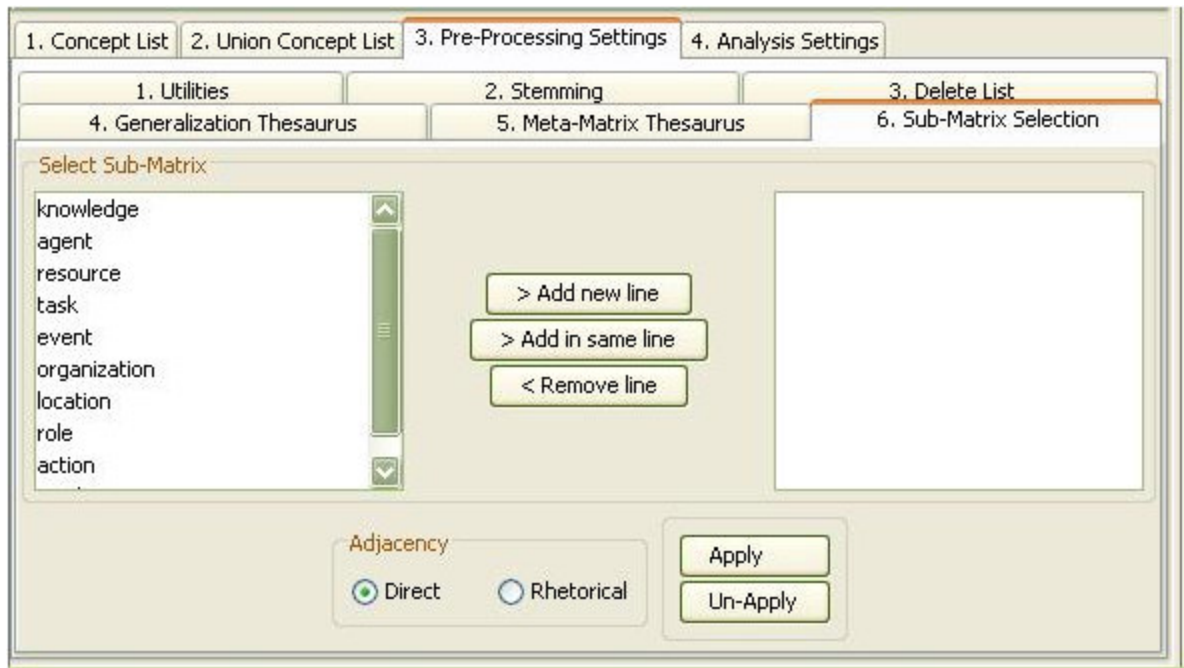
Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.



Then, open the Meta-Matrix Thesaurus by clicking the File menu, selecting Open Meta-Matrix Thesaurus, and choosing Open from highest level of pre-processing. The black ellipse in the screen shot below underscore how to open a Meta-Matrix Thesaurus from the file menu.



The union of concepts from the highest level of pre-processing will be displayed in alphabetical order on P2, tab no. 3 - Preprocessing Settings, tab no. 5 - Meta-Matrix Thesaurus. As we did not pre-process the text the original input sentence is used for input for the Meta-Matrix Thesaurus.



Furthermore, we have prepared a Meta-Matrix Thesaurus that we stored on our machine. This file looks like this:

- Mr\_Cray/agent
- Mrs\_Brown/agent
- Prof\_John\_Darren/agent
- dog/agent
- flowers/resource
- lotus/resource
- roses/resource
- carnations/resource
- daffodils/resource
- weeds/resource
- weeds/task
- planting/task
- eating/task
- yelling/task
- met/task

We clicked the File menu, selected Open Meta-Matrix Thesaurus and chose Open from file.

AutoMap searched the opened Meta-Matrix Thesaurus for the words contained in the prepared Thesaurus. When it found a match it assigned the words in the opened Thesaurus the Meta-Matrix categories that were assigned to the same concept in the pre-defined file. Below is the result:

1. Concept List 2. Union Concept List 3. Pre-Processing Settings 4. Analysis Settings

1. Utilities 2. Stemming 3. Delete List 4. Generalization Thesaurus 5. Meta-Matrix Thesaurus 6. Sub-Matrix Selection

Concept	knowledge	agent	resource	task	event	organi...	location	role
cray	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
cray's	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
craye	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
daffodils	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
darren	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
day	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dog	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dog!"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dug	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
eating	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
fence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
flowers	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

☐ Thesaurus content only

Adjacency  
☒ Direct ☐ Rhetorical

Apply  
Un-Apply

Now we edit the Thesaurus by modifying some of the pre-assignments (e.g., centre) and adding assignments for concepts not assigned to Meta-Matrix categories yet (e.g., contact, copenhagen). Not all concepts were associated with Meta-Matrix categories (e.g., mobile).

1. Concept List 2. Union Concept List 3. Pre-Processing Settings 4. Analysis Settings

1. Utilities 2. Stemming 3. Delete List 4. Generalization Thesaurus 5. Meta-Matrix Thesaurus 6. Sub-Matrix Selection

Concept	knowledge	agent	resource	task	event	organi...	location	role
cray	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
cray's	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
craye	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
daffodils	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
darren	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
day	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dog	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dog!"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dug	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
eating	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
fence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
flowers	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

☐ Thesaurus content only

Adjacency  
☒ Direct ☐ Rhetorical

Apply  
Un-Apply

In the next step we applied the Apply Meta-Matrix Thesaurus with the following settings:

### Setting

Not Thesaurus content only (default)

### Resulting text

mr . cray's brown < agent> < task> the < resource> blossom at 10 am . mrs . brown was unhappy with the < agent> . she < task> at it saying "you impossible dog!" but the < agent> kept < task> the < resource> and < task> . she asked mr .

cray to stop the < agent> . he couldn't . mrs . brown planted < resource> and weeded the garden . the silly < agent> % dug up the < resource> looking for a vole on june 12 , 1880 . weeding was no longer needed . prof . darren , mrs . brown & mr . cray < task> the next day to concoct a plan . < agent> and mrs . brown put up a scarecrow . she thought it would scare the < agent> . mr . craye put up a fence . problem solved . then mrs . brown planted < resource> , < resource> , < resource> , and < resource> .

Thesaurus content only, Direct Adjacency (default if Thesaurus content only is chosen)

. < agent> < task> < resource> . . < agent> . < task> < agent> < task> < resource> < task> . . < agent> . . . < resource> . < agent> < resource> . . . . < task> . < agent> . . < agent> . . . . < resource> < resource> < resource> < resource> .

Thesaurus content only, Rhetorical Adjacency

xxx . xxx xxx < agent> < task> xxx < resource> xxx xxx xxx  
xxx . xxx . xxx xxx xxx xxx xxx < agent> . xxx < task> xxx  
xxx xxx xxx xxx xxx xxx xxx < agent> xxx < task> xxx < resource> xxx < task> . xxx xxx xxx . xxx xxx xxx xxx < agent> . xxx xxx . xxx . xxx xxx < resource> xxx xxx xxx xxx . xxx xxx < agent> xxx xxx xxx xxx < resource> xxx xxx xxx xxx xxx xxx xxx xxx . xxx xxx xxx . xxx xxx . xxx xxx xxx . xxx < task> xxx xxx xxx xxx xxx xxx . < agent> xxx xxx . xxx xxx xxx xxx xxx . xxx xxx xxx xxx xxx xxx < agent> . xxx . xxx xxx xxx xxx . xxx xxx . xxx xxx . xxx xxx < resource> < resource> < resource> xxx < resource> .

## Text PreProcessing

### 8. Sub-Matrix Selection

The Sub-Matrix Selection enables the user to re-translate concepts represented by a Meta-Matrix category in order to run Sub-Matrix Analysis. If input texts (no matter if they were pre-processed with a Delete list or not) were used in order to generate the Concept List for the Meta-Matrix Thesaurus, concepts represented by a Meta-Matrix category will be translated into text-level concepts. If input texts (no matter if they were pre-processed with a Delete list or not) were pre-processed with a Generalization Thesaurus before applying the Meta-Matrix Thesaurus, concepts represented by a Meta-Matrix category will be translated into key concepts.

The Thesaurus content only option always automatically applies for the Sub-Matrix Selection.

## ***8.1 Select Sub-Matrix Categories/***

**Precondition: Sub-Matrix Selection can only be performed if Meta-Matrix Thesaurus was applied. There are 4 ways to select sub matrices:**

### ***8.1.1 If you do not have a sub matrix selection file available***

Create and modify a sub matrix selection.

### ***8.1.2 If you do not have a sub matrix selection file available and want to select the full meta matrix (means all cells in the meta matrix)***

Go to the File menu, Select Open Sub-Matrix Selection, click on Select Full Meta Matrix. AutoMap dynamically generates all combinations of meta matrix categories as specified in the meta matrix thesaurus, including user-defined categories, that represent all cells of the meta matrix and display these combinations in the left window on P2, 8. Sub-Matrix Selection.

You can modify this sub matrix selection.

This is the full meta-matrix:

- agent/agent/knowledge/organization/task-event/resource/location/role/action/attribute
- knowledge/agent/knowledge/organization/task-event/resource/location/role/action/attribute
- organization/agent/knowledge/organization/task-event/resource/location/role/action/attribute
- task-event/agent/knowledge/organization/task-event/resource/location/role/action/attribute
- resource/agent/knowledge/organization/task-event/resource/location/role/action/attribute
- location/agent/knowledge/organization/task-event/resource/location/role/action/attribute
- role/agent/knowledge/organization/task-event/resource/location/role/action/attribute
- action/agent/knowledge/organization/task-event/resource/location/role/action/attribute
- attribute/agent/knowledge/organization/task-event/resource/location/role/action/attribute

### ***8.1.3 If you have a sub matrix selection file available***

Go to the File menu, Select Open Sub-Matrix Selection, click on Open from file.

You can modify this sub matrix selections.

#### ***8.1.4 If you want to write your own sub matrix selection file and load it into AutoMap:***

Build your own sub matrix selection file outside of AutoMap, save it in .txt format, and load it into AutoMap.

Go to the File menu, Select Open Sub-Matrix Selection, click on Open from file. You can modify this sub matrix selections.

### ***8.2 Create or Modify Sub-Matrix Selection***

Go to P2, 8. Sub-Matrix Selection, click on a Sub-Matrix category you wish to select and hit the > Add in same line button. The selected category appears in the right text field on P2, (tab no. 8) Sub-Matrix Selection.

You can select as many Sub-Matrix categories per row as you wish by clicking on a category in the left window on P2, tab no. 8 Sub-Matrix Selection and move it to the right window by clicking the > Add in same line button. Additionally, you can select as many rows of sub matrix selections as you wish by clicking on a category in the left window on P2, tab no. 8 Sub-Matrix Selection and move it to the right window by clicking the > Add in new line button. To add a further category to a new line, first single click on this category in the right window on P2, tab no. 8 Sub-Matrix Selection, so that the category is highlighted in blue. To unselect a selected Sub-Matrix category, click on the row in the right window and hit the < Remove line button on P2, tab no. 8 Sub-Matrix Selection. The row will disappear from the right window.

#### **Tip!**

- **If you wish to analyze any relation between e.g. agents and organizations, you need to select agent/ organization in one row, and organization/ agent in another row. If you wish to analyze any relations among all agents, select agent/agent.**
- **If you wish to analyze any relations among all agents and their relation with all organizations, select agent/agent/organization and in another row organization/agent/agent/.**

### ***8.3 Apply Sub-Matrix Selection***

1. The Thesaurus content only option always automatically applies for the Sub-Matrix Selection. Select an adjacency option. Adjacency can be either direct (default) or rhetorical.  
If you do not change the adjacency option, direct adjacency will be applied.
2. To apply your Meta-Matrix Thesaurus with the settings you have specified click the Apply Sub-Matrix Selection button on the Sub-Matrix Selection Index

card.

AutoMap uses the entries in the Thesaurus to search the text(s) for meta-matrix tags. If a match is found it will be translated into the related concept. The Thesaurus IS case insensitive.

3. See the pre-processed texts on the P1, tab no. 6 Sub-Matrix Text.

### ***8.3.1 Direct Adjacency***

Direct adjacency means that original distances of concepts that represent meta-matrix categories will neither be visualized nor considered for analysis.

To choose the direct adjacency click the Direct button in the Adjacency field on P2, tab no. 8 Sub-Matrix Selection index card. Then apply the Sub-Matrix Selection.

If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.

As a result, only concepts that represent meta-matrix categories are displayed on P1, tab no. 6 Sub-Matrix Text index card and will be considered for analysis. All concepts in the resulting text appear directly adjacent to each other.

### ***8.3.2 Rhetorical Adjacency***

To choose the rhetorical adjacency click the Rhetorical button in the Adjacency field on P2, tab no. 8 Sub-Matrix Selection index card. Then apply the Sub-Matrix Selection. If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.

As a result, concepts that represent meta-matrix categories are displayed on P1, tab no. 6 Sub-Matrix Text and will be considered for analysis. Original distances of concepts that represent meta-matrix categories will be visually symbolized by placeholders (xxx) and considered for analysis.

## ***8.4 Un-Apply a Sub-Matrix Selection***

To un-apply a Sub-Matrix Selection that was applied to the data, go to P2, tab no. 8 Sub-Matrix Selection index card and hit the Un-Apply button. The tab no. 6 Sub-Matrix Text on P1 will be cleared.

## ***8.5 Save Sub-Matrix Selection***

Apply the Sub-Matrix Selection before you save it.

To save a Sub-Matrix Selection (the content of the right window on P2, tab no. 8 Sub-Matrix Selection, click the File menu, select Save applied Sub-Matrix Selection as applied.

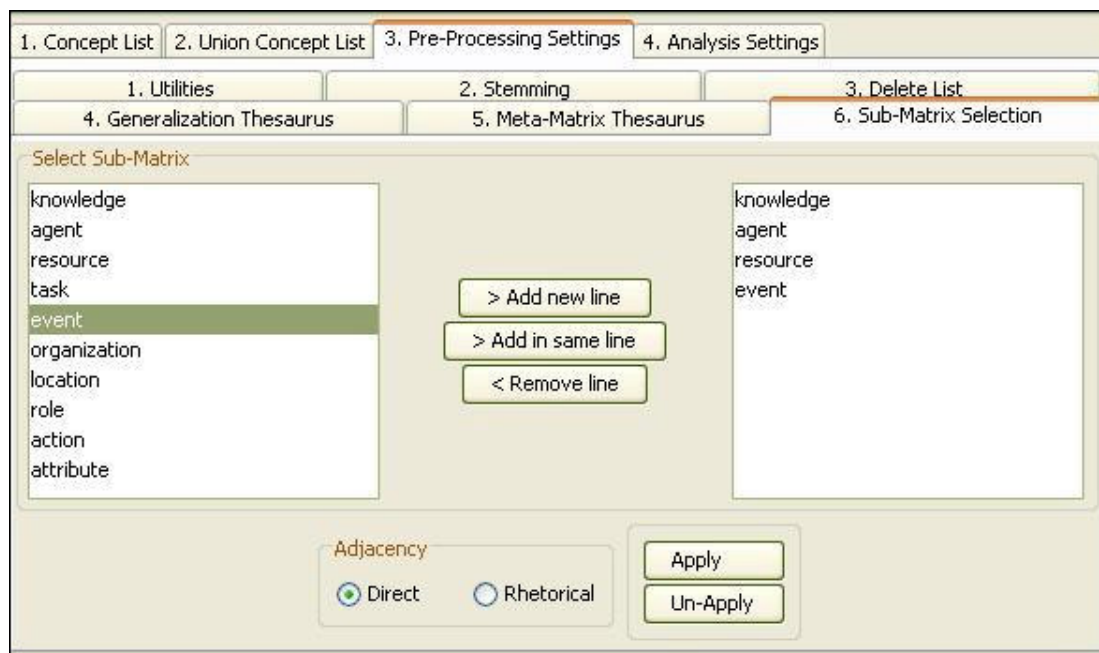
## 8.6 Save text(s) after Sub-Matrix Selection

To save the text(s) after the application of the Sub-Matrix Selection, click the File menu, select Save Text(s) after Sub-Matrix Selection applied. All texts are automatically saved in a folder called "preprocessed" in the root directory of AutoMap. The filename will be after\_SubMatrixSelection\_NameOfYourText.txt".

## 8.7 Example for Sub-Matrix Selection

The Sub-Matrix Selection as shown here is based on the example for the Meta-Matrix Thesaurus.

We opened P2, tab no. 3 - Pre-Processing Settings, tab no. 6- Sub-Matrix Selection.



Then we applied the Sub-Matrix Selection with the following settings:

Setting	Resulting text
Select:	
Agent/ Location/ Action	
Location/ Agent/ Action	. dog.. dog. dog.. dog.... dog,.. ,,... prof_john_darren..
Action/ Location/ Agent	dog.....,
Direct Adjacency (default)	
Select:	xxx. xxx xxx dog xxx xxx xxx xxx xxx xxx. xxx. xxx
Agent/ Location/ Action	xxx xxx xxx xxx dog. xxx xxx xxx xxx xxx xxx xxx
Location/ Agent/ Action	xxx xxx dog xxx xxx xxx xxx xxx. xxx xxx xxx. xxx
Action/ Location/ Agent	xxx xxx xxx dog. xxx xxx. xxx. xxx xxx xxx xxx xxx
Rhetorical Adjacency	xxx. xxx xxx dog xxx xxx xxx xxx xxx xxx xxx xxx



(see also next picture) xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx

xxx

xxx. xxx, xxx. xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx  
 xxx xxx. prof\_john\_darren xxx xxx. xxx xxx xxx xxx xxx.  
 xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx  
 xxx. xxx xxx. xxx xxx xxx, xxx, xxx, xxx xxx.

Select:  
 Full Meta-matrix  
 Direct Adjacency  
 (default)

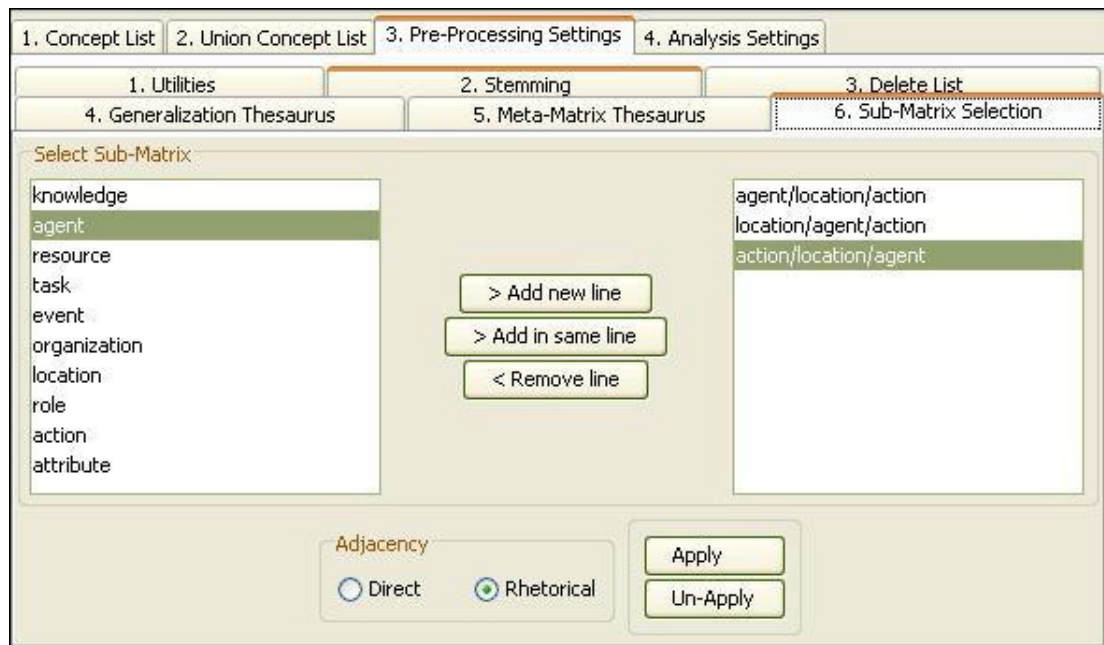
. dog eating lotus.. dog. yelling dog eating flowers  
 weeds.. dog... roses. dog roses,.. ,... met.  
 prof\_john\_darren.. dog..... lotus, carnations, daffodils,  
 roses.  
 xxx. xxx xxx dog eating xxx lotus xxx xxx xxx xxx. xxx.  
 xxx xxx xxx xxx xxx dog. xxx yelling xxx xxx xxx xxx xxx  
 xxx xxx xxx dog xxx eating xxx flowers xxx weeds. xxx  
 xxx xxx. xxx xxx xxx xxx dog. xxx xxx. xxx. xxx xxx roses  
 xxx xxx xxx xxx. xxx xxx dog xxx xxx xxx xxx roses xxx  
 xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx

Select:  
 Full Meta-matrix  
 Rhetorical Adjacency  
 (default)

xxx

xxx. xxx, xxx. xxx xxx xxx. xxx met xxx xxx xxx xxx xxx  
 xxx xxx. prof\_john\_darren xxx xxx. xxx xxx xxx xxx xxx.  
 xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx  
 xxx. xxx xxx. xxx xxx lotus, carnations, daffodils, xxx  
 roses.

Detailed example for Sub-Matrix Selection:



# Statement Formation Choices

Statement formation choices can be made after pre-processing data and before running analysis. These choices define if, how, and where concepts will be linked. Statement formation settings will be applied to the highest level of data pre-processing that was applied. If no pre-processing was performed, statement formation settings apply to the original input text. If the user does not modify the statement formation settings AutoMap uses a set of default settings.

## 1. Analysis Settings

To specify the Analysis Settings or make the Analysis Settings, use the Analysis Settings Index Card.

Your settings will be automatically applied in the analysis. You do not need to confirm them.

If you do not want to change any of the suggested options a set of standard settings will be applied. The screen shot below shows an example of the Automap Analysis Settings tab in P2.

1. Concept List 2. Union Concept List 3. Pre-Processing Settings 4. Analysis Settings

1. Analysis Settings 2. Output Options

**Coding Ties**

Directionality Strength

☒ Uni-directional ☒ Frequency

☐ Bi-directional

**Windowing**

Punctuation Window-Size

☒ Ignore punctuation completely 2

☐ Reset window at the end of paragraphs only Find size of largest text in text set

☐ Reset window at the end of paragraphs and sentences

## Overview on the possible Settings:

### Coding Ties

Specify the way statements are counted.

#### Directionality

Select one of the following two possibilities by checking the button.

#### Strength

Strength will be printed

Uni-Directional

(When coding a tie, only 1st->2nd concept should be noted)

Bi-Directional (When coding a tie, both 1st <-> 2nd concept shall be noted)

**Frequency** (The cumulative frequency of every existing statement.)

out by default. To not print out Strength uncheck the item.

### Windowing

Windowing is a method that codes the text as a map by placing relationships between pairs of Concepts that occur within a window.

### Punctuation

Select one option by checking the radio button.

### Window Size

Select one window size by using the number chooser.

**Item not checked:** Existence of frequency will be printed out (binary result).

### Ignore punctuation completely

(Statements will be placed between all concepts.)

**Reset window at end of paragraphs only** (Statements will be placed only within every single paragraph.)

**Reset window at end of paragraphs and sentences** (Statements will be placed only within every single sentence.)

Window size between **2 and 100**. The Window Size defines how distant concepts can be and still have a relationship. Only concepts in same window can form statements.

## 1.1 Default Settings

If you do not want to change any of the suggested options the analysis will be done with the following Standard Settings:

The screenshot shows the '4. Analysis Settings' tab in the AutoMap software. It contains two sub-tabs: '1. Analysis Settings' and '2. Output Options'. The '1. Analysis Settings' sub-tab is active. It is divided into two main sections: 'Coding Ties' and 'Windowing'. In the 'Coding Ties' section, under 'Directionality', the 'Uni-directional' radio button is selected, and under 'Strength', the 'Frequency' checkbox is unchecked. In the 'Windowing' section, under 'Punctuation', the 'Ignore punctuation completely' radio button is selected. Under 'Window-Size', a dropdown menu shows the value '2', and there is a button labeled 'Find size of largest text in text set'.

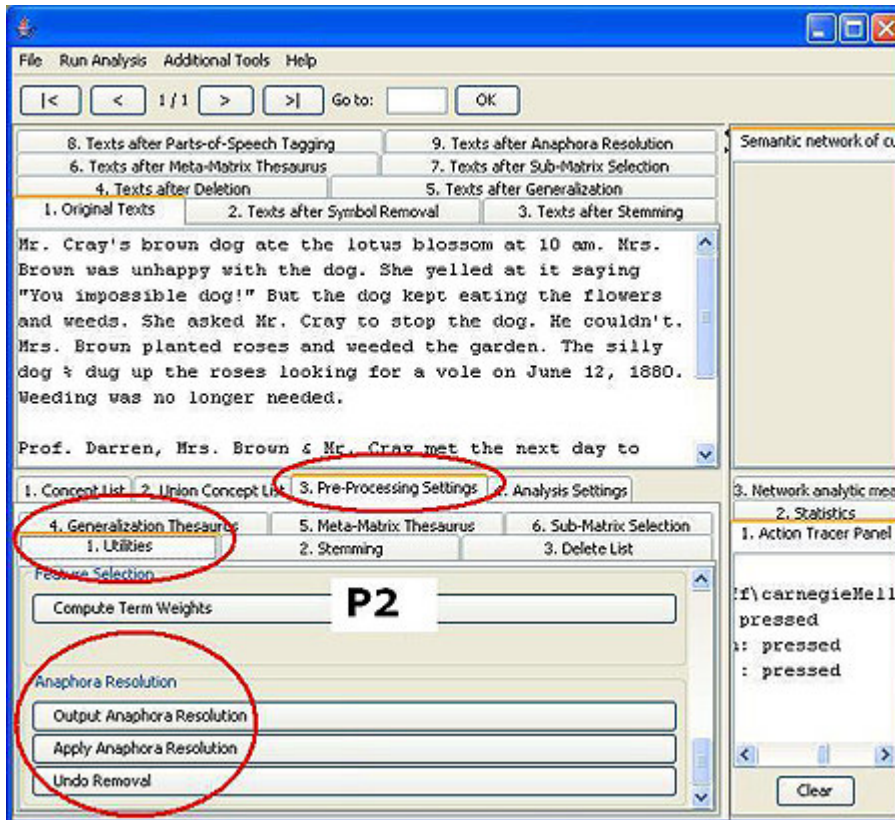
## Anaphora Resolution

Automap can perform anaphora resolution.

Anaphora is a linguistic instance of a grammatical expression referring to another. In a general linguistic sense, an anaphoric expression is one represented by some kind of

deictic, a process whereby words or expressions rely absolutely on context. Sometimes this context needs to be identified.

The anaphora resolution function is a pre-processing setting and can found in Panel P2, Tab no. 3. Pre-processing settings in Automap's main interface. The screen shot below shows where the tool is located.



This tool extracts network information from text that requires additional textual references from within a certain phrase or grammatical construct.

To Access Anaphora Resolution: Panel P2 > Tab no. 3 Pre-Processing Settings > 1. Utilities > scroll down to bottom of window

Some examples of what would constitute an Anaphora Resolution might include:

- The boy took the hamburger and ate it. It is anaphoric under the strict definition (it refers to the hamburger).
- Larry called off work because he felt sick. he is anaphoric (it refers to Larry).

Moreover, think of a sentence being taken out of context and the missing information required to fully explain the meaning.

*Example: The President of our company came to visit us yesterday. The visit was the first time he had come to Pittsburgh since 1998.*

If the second sentence stands alone by itself, it is necessary to resolve the anaphor.

Automap can apply this pre-process settings, identify Anaphora Resolution instances and then have them isolated and removed, which is activated by selecting the first and middle button under the Anaphora Resolution tool. The third button allows you to output the Anaphora Resolution contained in your text. Each of these preprocessing settings can be undone.

## Statement Formation Choices

Statement formation choices can be made after pre-processing data and before running analysis. These choices define if, how, and where concepts will be linked. Statement formation settings will be applied to the highest level of data pre-processing that was applied. If no pre-processing was performed, statement formation settings apply to the original input text. If the user does not modify the statement formation settings AutoMap uses a set of default settings.

### 1. Analysis Settings

To specify the Analysis Settings or make the Analysis Settings, use the Analysis Settings Index Card.

Your settings will be automatically applied in the analysis. You do not need to confirm them.

If you do not want to change any of the suggested options a set of standard settings will be applied. The screen shot below shows an example of the Automap Analysis Settings tab in P2.

### Overview on the possible Settings:

#### Coding Ties

Specify the way statements are counted.

#### Directionality

Select one of the following two possibilities by checking the button.

#### Strength

Uni-Directional

(When coding a tie, only 1st->2nd concept should be noted)

Bi-Directional (When coding a tie, both 1st <-> 2nd concept shall be noted)

**Frequency** (The cumulative frequency of



Strength will be printed out by default. To not print out Strength uncheck the item.

### Windowing

Windowing is a method that codes the text as a map by placing relationships between pairs of Concepts that occur within a window.

### Punctuation

Select one option by checking the radio button.

### Window Size

Select one window size by using the number chooser.

every existing statement.)

**Item not checked:** Existence of frequency will be printed out (binary result).

### Ignore punctuation completely

(Statements will be placed between all concepts.)

**Reset window at end of paragraphs only** (Statements will be placed only within every single paragraph.)

**Reset window at end of paragraphs and sentences** (Statements will be placed only within every single sentence.)

Window size between **2 and 100**.

The Window Size defines how distant concepts can be and still have a relationship. Only concepts in same window can form statements.

## 1.1 Default Settings

If you do not want to change any of the suggested options the analysis will be done with the following Standard Settings:

## Semantic Text Analysis

Automap contains text analysis utilities to help you in the pre-processing and data analysis of your text examples. Take time to become familiar with them as they allow you to quickly work with your text examples.

## 1. Browse through texts

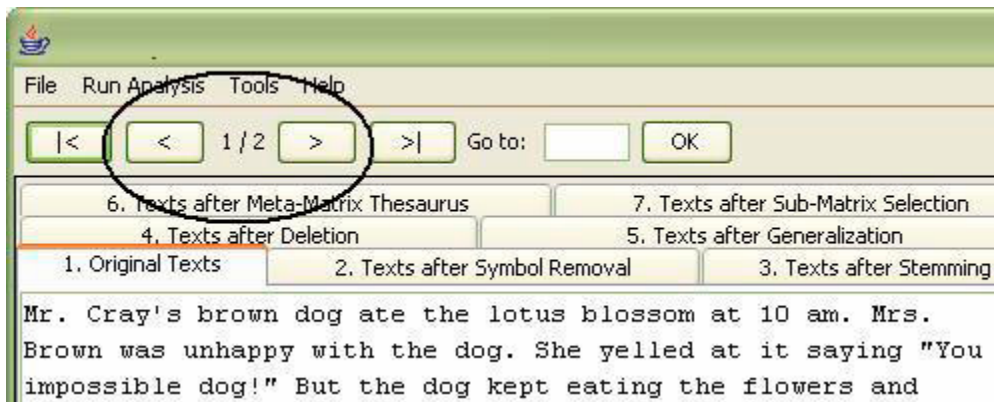
This function enables you to quickly jump from text example to another. All panels are synchronized in the Browse Menu. A series of screen shots below the instruction ad emphasis on how to access text Automap's text browsing features.

How to use the Browse Menu:

To go backward or forward text by text:

**Click the ">" button or the "<" button in the browse menu bar.**

To go to first or last text in the text set:



To go to first or last text in the text set:

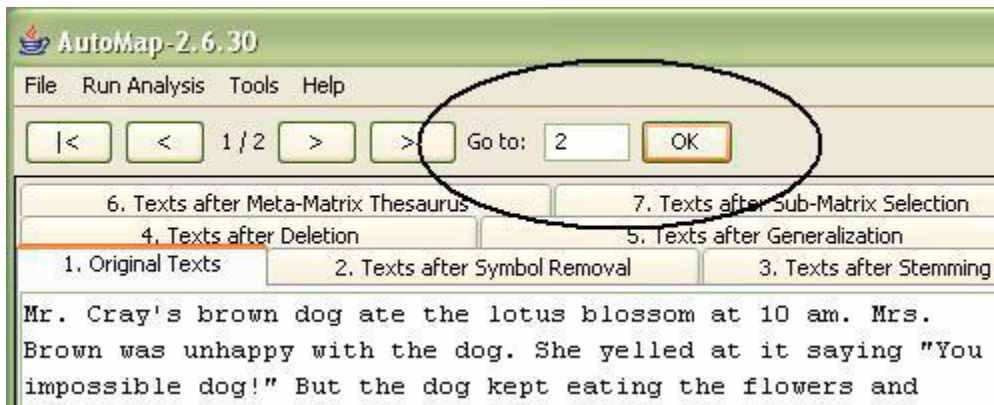
**Click the ">>" button or the "<<" button.**



To go to a specific text:



**Enter the text number in text field right next to the Go to command and hit OK.**

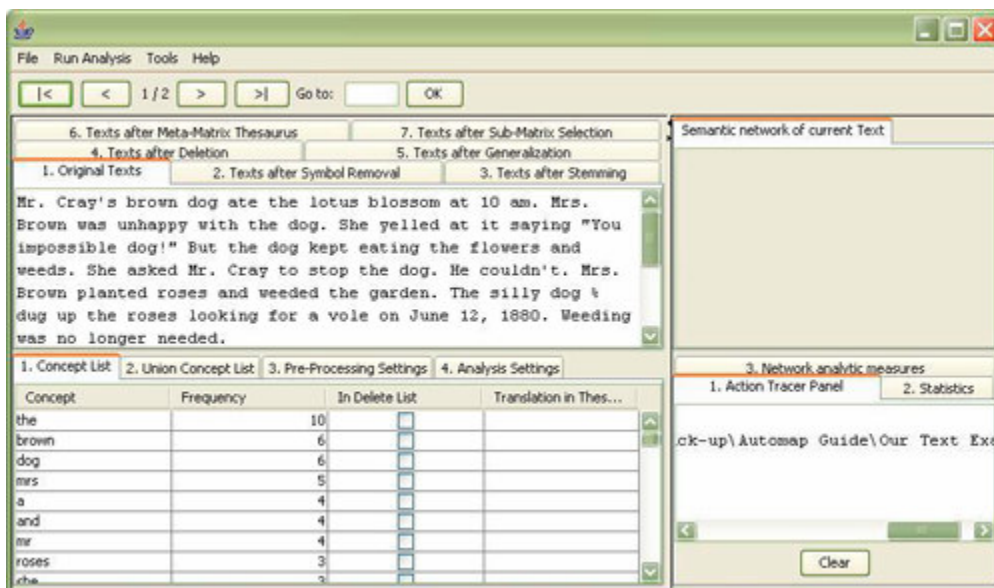


The name of the currently selected text is displayed on the Browse menu.

These files can be browsed:

- All input texts (P1, No. 1. Input Text index card).
- All texts after each stage of pre-processing (all index cards of P1).
- Concept lists per text (original input text or texts after each stage of pre-processing).
- Map (P3) and Statistics (P4) outputs that relate to the text currently displayed on the index cards of P1.

## 1.1 Example for Browse Menu



The example shows a part of the "Our Text 1.txt" in panel P1 tab no. 1 Original Texts.

The browse menu tells you several important facts:

- The data set opened contains two texts.
- The first of the loaded texts is displayed.
- The filename of the displayed text.

## ***2. Concept List***

The Concept List is displayed in panel P2 tab No. 1 Concept List. The Concept List is created automatically once a text or a set of texts are loaded or modified in Automap.

The concept list tells you several important facts about your text:

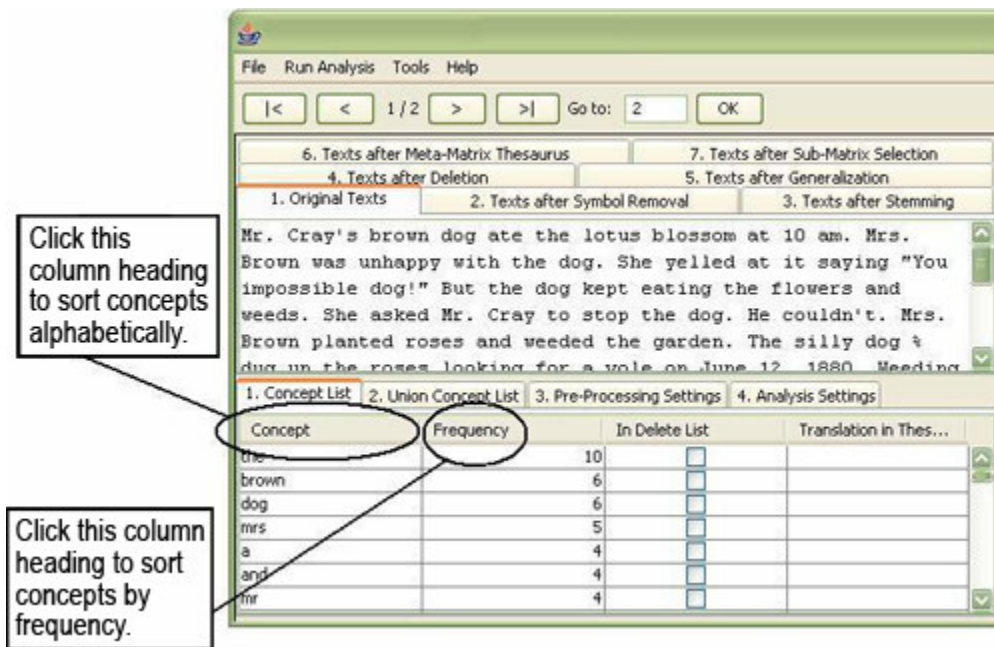
- 
- The number of concepts found in the text displayed in P1 tab No. 1 Original Text.
- Related frequencies of those concepts.
- Cumulated unique concepts and total concepts contained in the data set.

**Tip! The number of unique concepts considers each concept only once, whereas the number of total concepts considers repetitions of concepts.**

By default, the Concept List is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. In order to resort the list, click on the header of the second-column header Frequency.

### ***2.1 Example for Concept List***

The example below shows a part of the Concept List for the text displayed in the browse menu. The Table is ordered by Frequency. The concept list contains more entries than the interface can display:



### 3. Create and refresh Union Concept List

The Union Concept List, found in panel P2, differs from the Concept List (tab no. 1) in that it considers concepts across all texts loaded in Automap, rather than one single text file. There are several key pieces of information the Union Concept List tells you:

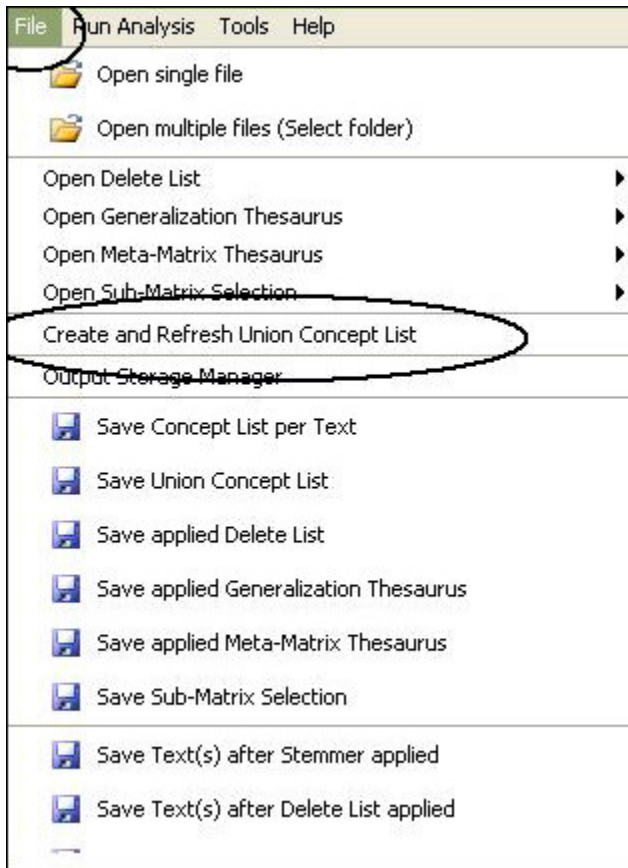
- Concepts contained in all loaded text sets.
- Related, cumulative frequencies of concepts in all text sets.
- Cumulative unique and total concepts.

Union Concept List results are displayed on tab No. 2 Union Concept List in P2. However, you must first refresh the union concept list from the file menu, before viewing your results on the No. 2 Union Concept tab. The union concept list can be refreshed after each step of pre-processing in order to visualize the impact of pre-processing operations on the union of concepts.

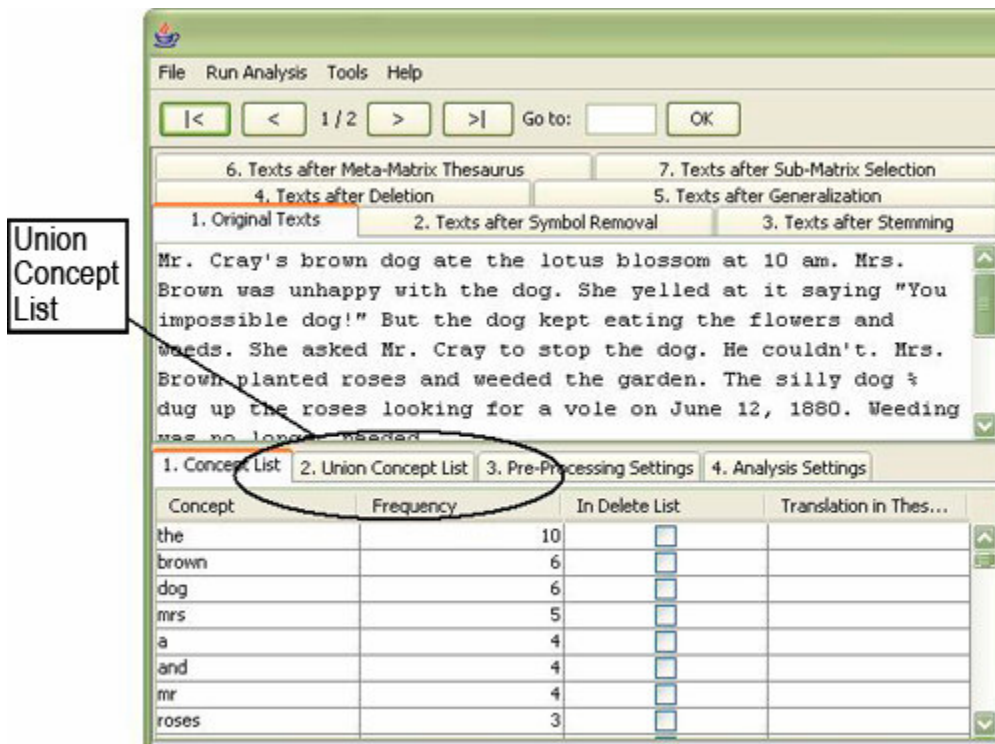
To refresh the Union Concept, from the drop-down menu bar:

#### **File menu > Refresh Union Concept List.**

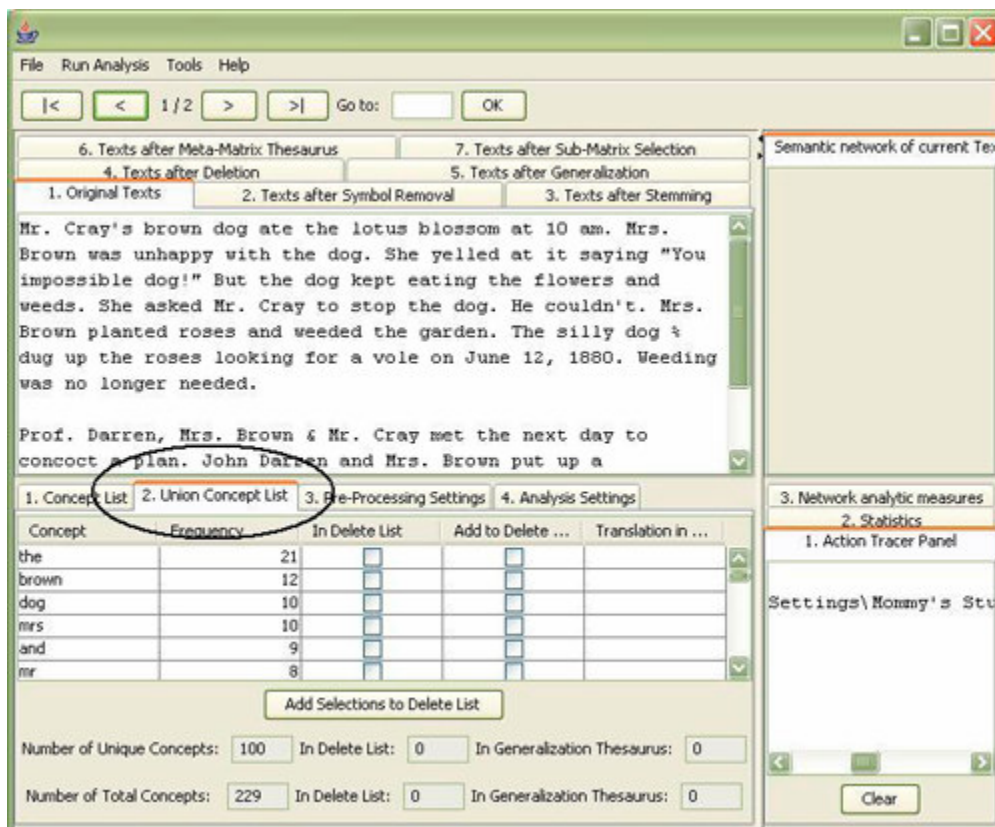
The ellipse in the screen shot below shows where to access the Refresh Union Concept List from the drop-down menu bar:



The call out box in the screen shot below, shows where to locate the Union Concept tab, which will contain the results of the Refresh Union Concept List analysis:



Your results will be displayed in tab No 2. Union Concept List. The black ellipse in the screen shot below highlights where to find this tab in the Automap GUI.



By default, the list is sorted by decreasing frequency of concepts. In order to sort the list alphabetically click on the first-column header Concept. To re-sort the list, click on the second-header column Frequency.

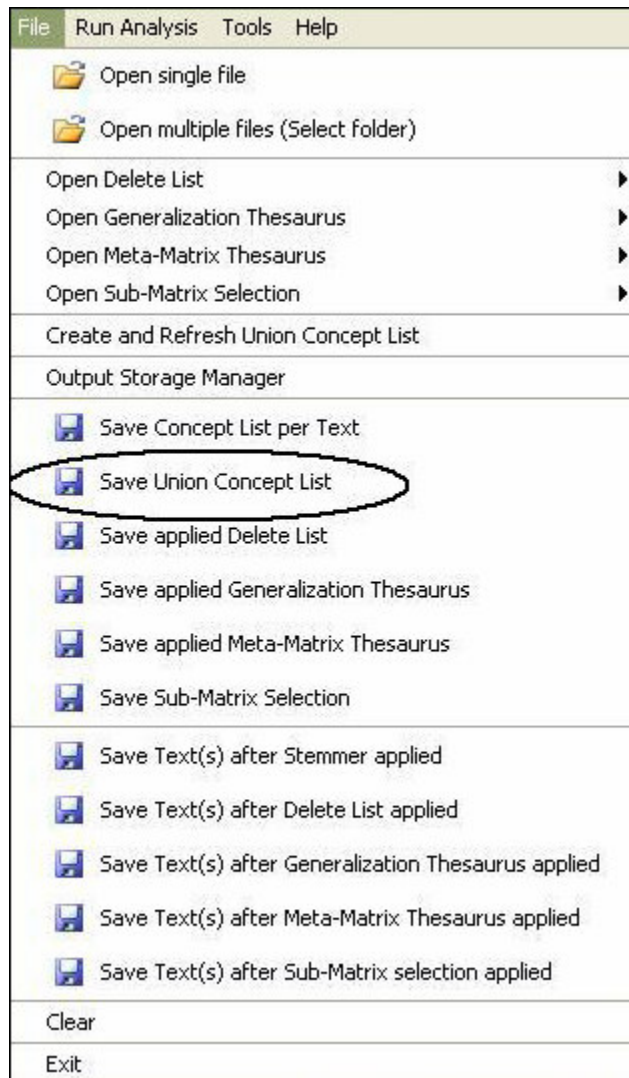
**Note: The number of unique concepts considers each concept only once, whereas the number of total concepts also considers repetitions of concepts.**

### ***3.1 Save Union Concept List***

To save a Union Concept List follow these steps:

**File menu > Save Union Concept List.**

A file chooser will pop up. The black ellipse below highlights how to save a Union Concept List from the drop-down menu bar.





### **3.2 Example for Union Concept List**

Let us walk through an example of creating a Union Concept List working with our text examples from above. They are restated below for you convenience:

1) Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed.

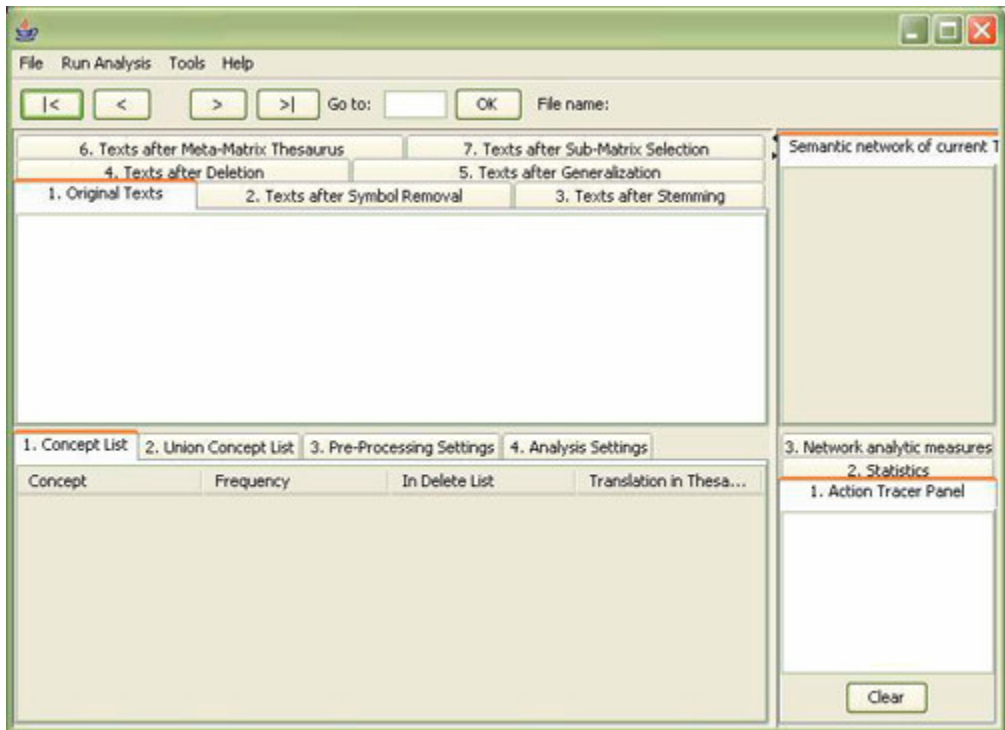
Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

2) Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray was pleased too.

Mrs. Brown watered the roses and fertilized the garden on June 13, 1880.

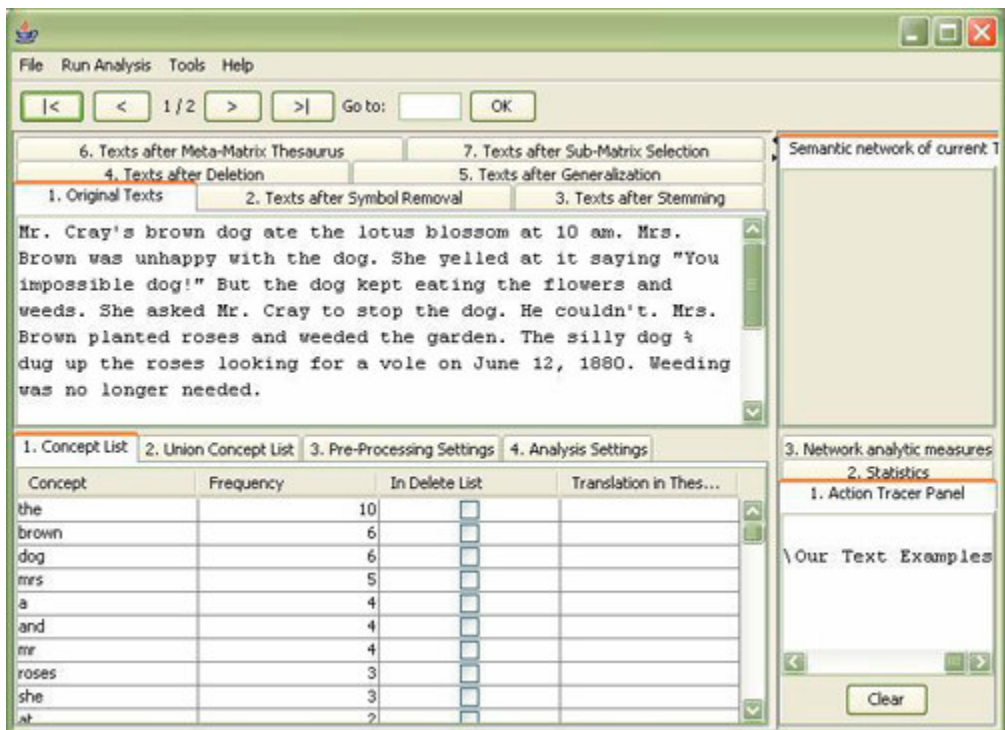
Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed how the plan had worked. John Darren and Mrs. Brown would take down the scarecrow the following week. She thought it was too scary for the dog. Mr. Craye painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.

The first step is to load in your multiple texts from which we will create and save a Union Concept List. An empty AutoMap GUI is shown below before any text is loaded into it. This is what you will see when you first run AutoMap:



From the drop-down menu:

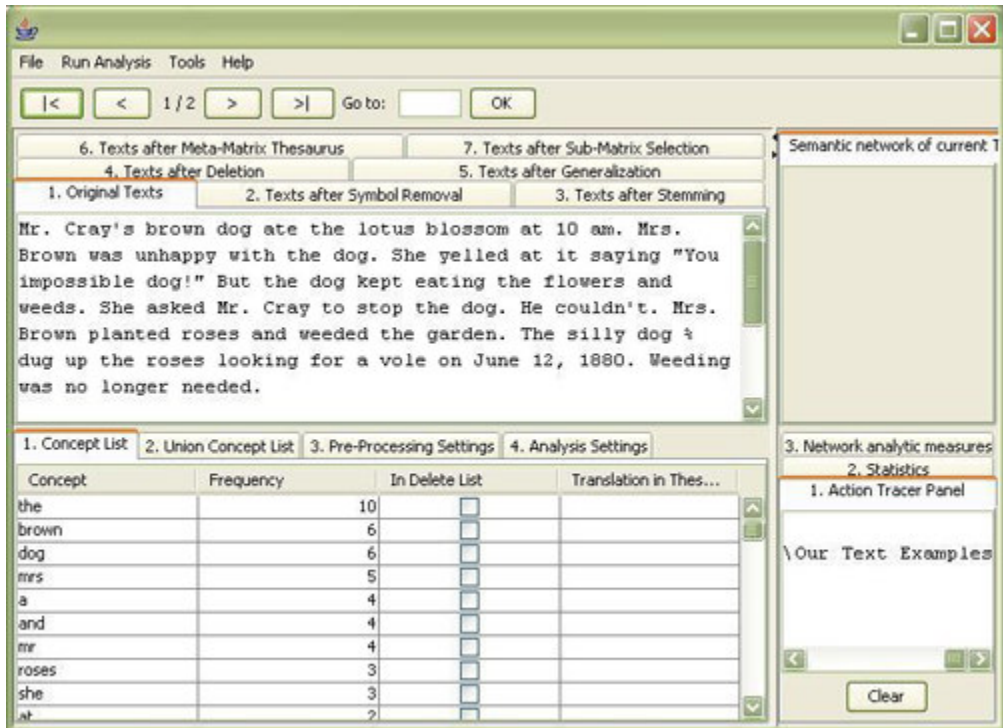
**File > Open Multiple Files > (select location of folder on applicable drive containing your text examples)**





## Select "Open"

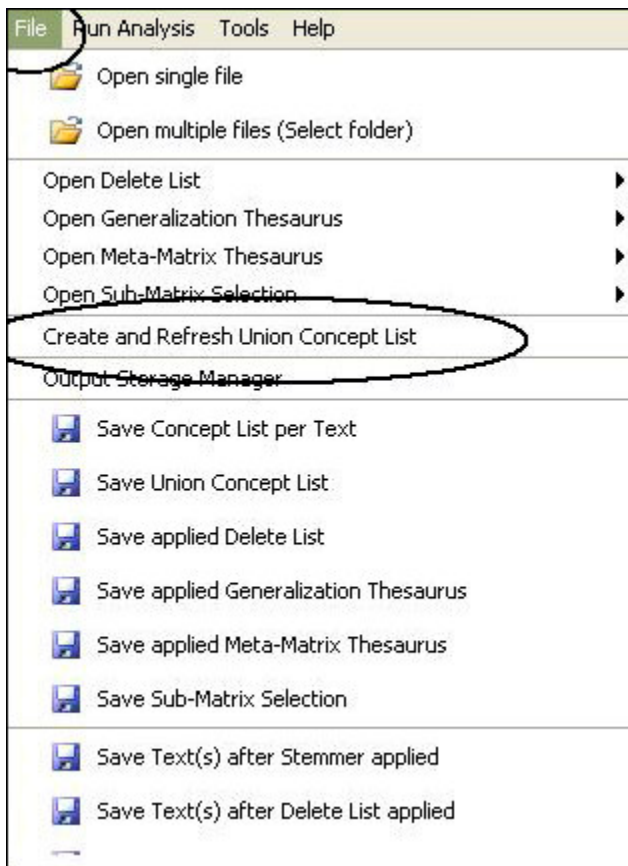
The screen shot below displays the Automap GUI after our text examples have been loaded into it.



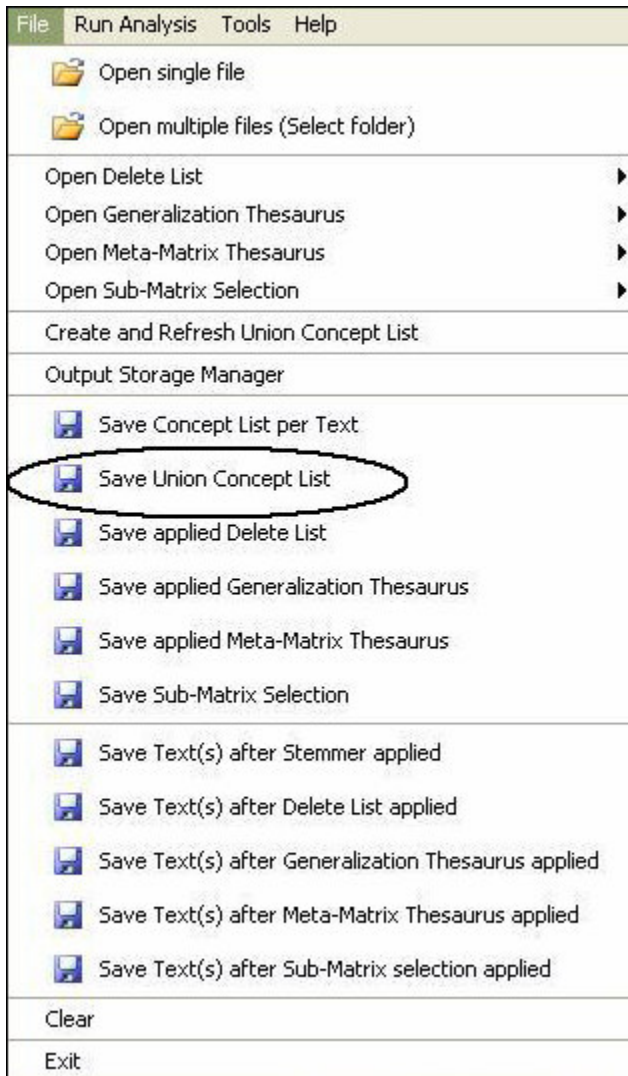
To run the Create and Refresh Union Concept Utility based on multiple texts:

## File > Create and Refresh Union Concept List

The following series of screen shots present a step-by-step on how to create an refresh a union concept list:



Select "Save Union Concept List" from the drop-down menu:



The result is shown in the screen shot below:

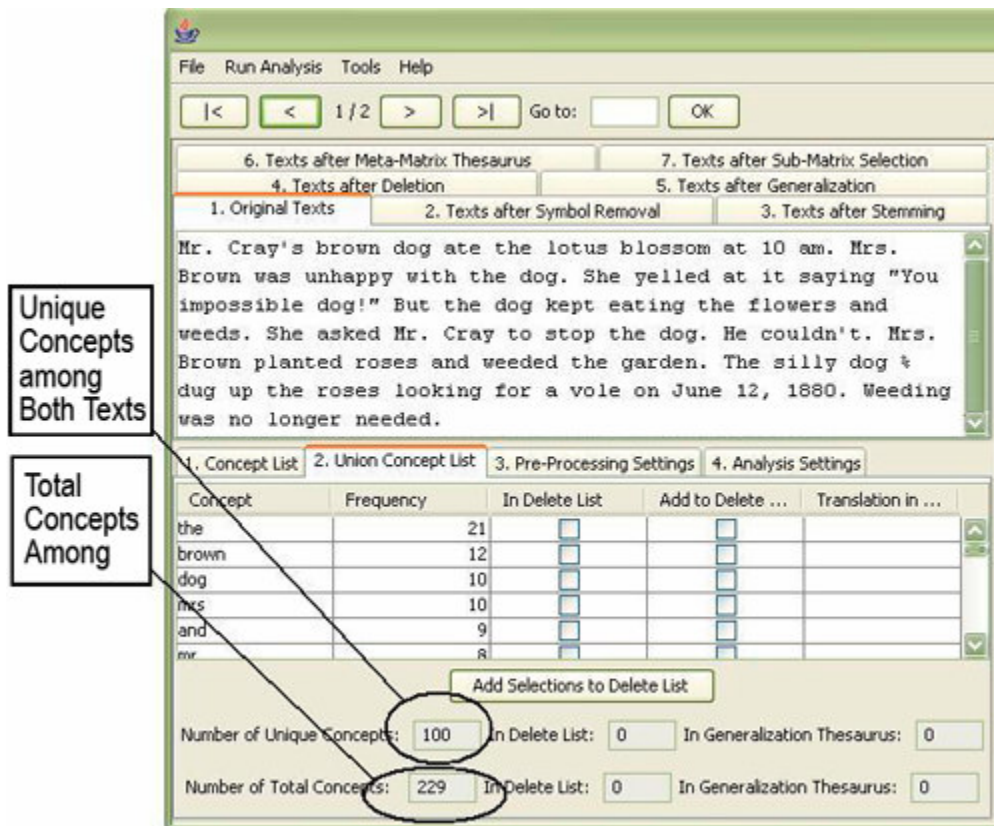
The screenshot shows the AutoMap software interface. At the top, there's a menu bar with 'File', 'Run Analysis', 'Tools', and 'Help'. Below it are navigation buttons: '<|', '<', '1 / 2', '>', '>|', 'Go to:', and 'OK'. A series of tabs represent different processing stages: 1. Original Texts, 2. Texts after Symbol Removal, 3. Texts after Stemming, 4. Texts after Deletion, 5. Texts after Generalization, 6. Texts after Meta-Matrix Thesaurus, and 7. Texts after Sub-Matrix Selection. The main text area displays two paragraphs of sample text. Below the text is a tabbed interface with '1. Concept List', '2. Union Concept List' (which is circled), '3. Pre-Processing Settings', and '4. Analysis Settings'. The 'Union Concept List' tab contains a table of concepts and their frequencies. Below the table is an 'Add Selections to Delete List' button. At the bottom, there are input fields for 'Number of Unique Concepts' (100), 'In Delete List' (0), and 'In Generalization Thesaurus' (0), along with their total counts. On the right side, there are three panels: 'Semantic network of current T', '3. Network analytic measures', and '2. Statistics', with a 'Clear' button at the bottom.

Concept	Frequency	In Delete List	Add to Delete ...	Translation in ...
the	21	<input type="checkbox"/>	<input type="checkbox"/>	
brown	12	<input type="checkbox"/>	<input type="checkbox"/>	
dog	10	<input type="checkbox"/>	<input type="checkbox"/>	
mrs	10	<input type="checkbox"/>	<input type="checkbox"/>	
and	9	<input type="checkbox"/>	<input type="checkbox"/>	
me	8	<input type="checkbox"/>	<input type="checkbox"/>	

Number of Unique Concepts: 100 In Delete List: 0 In Generalization Thesaurus: 0

Number of Total Concepts: 229 In Delete List: 0 In Generalization Thesaurus: 0

The table is ordered by Frequency. The concept list contains more entries than the interface can display. The displayed Union Concept List indicates there are 100 unique concepts and 229 total concepts in the data set.



## 4. Semantic Network Analysis

Semantic Network Analysis can be run on original Input text(s) or texts that have been pre-processed with a Delete List and/ or a Generalization thesaurus.

Before you run map analysis make sure that you have completed the following steps:

- Pre-processed the texts if you wish to.
- Specified the analysis settings. If you do not do so, the default setting default settings will be applied.
- Specified additional output options if you wish to. If you do not select additional output options, the default outputs default outputs will be generated

If you wish to analyze a single text, click the Run Analysis menu and select Single Map Analysis.

If you wish to analyze a set of texts, click the Run Analysis menu and select Multiple Map Analysis.

The Results will be displayed on P3 on the Map index card and on P4 on the Stat index card. If you had requested additional outputs, those will be generated and stored under the directories specified under the Section Additional Outputs.

## Other Semantic Network Analysis Points to Consider

- If you have analyzed multiple texts, you can browse through the results and see the related texts.  
Results of multiple analyses are automatically saved in folder named "output" under directory where AutoMap 2.0 is installed.
- This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as a stat\_output.xls file that contains the stats of all texts.
- The "Stat Output" folder is overwritten with every new analysis you run. So if you want to save the results of a current "Stat Output" folder just rename the folder.

### 4.1 Example for Semantic Network Analysis

An extract from the Denmark text was used as input:

Reporters said hundreds of people emerged from shops in Copenhagen city centre to see what was happening, and used their mobile phones to contact their families.

The text was pre-processed with AutoMap's customized extensive Delete List. These are the resulting texts:

Input text	Tool used	Setting	Resulting text
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted	AutoMap's customized extensive Delete List:		
a			
and			
as			
at			
but			
for			
he			
her			
her			
hers			
him			
his			
i			
it			
its			
me			
mine			
my			
nor			
		Direct Adjacency (default)	Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding was no longer needed.
			Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses..
		Rhetorical Adjacency	Mr. Cray's brown dog ate xxx lotus blossom xxx 10 am. Mrs. Brown was unhappy with xxx

roses and of  
 weeded the or  
 garden. The our  
 silly dog % she  
 dug up the so  
 roses that  
 looking for a the  
 vole on June their  
 12, 1880. theirs  
 Weeding them  
 was no they  
 longer to  
 needed. us  
 we  
 Prof. who  
 Darren, Mrs. whoever  
 Brown & whom  
 Mr. Cray whomever  
 met the next will  
 day to would  
 concoct a you  
 plan. John your  
 Darren and yours  
 Mrs. Brown yourself  
 put up a  
 scarecrow.  
 She thought  
 it would  
 scare the  
 dog. Mr.  
 Craye put up  
 a fence.  
 Problem  
 solved. Then  
 Mrs. Brown  
 planted  
 lotus,  
 carnations,  
 daffodils,  
 and roses.

dog. xxx yelled xxx xxx saying  
 "You impossible dog!" xxx xxx  
 dog kept eating xxx flowers xxx  
 weeds. xxx asked Mr. Cray xxx  
 stop xxx dog. xxx couldn't. Mrs.  
 Brown planted roses xxx  
 weeded xxx garden. xxx silly  
 dog % dug up xxx roses looking  
 xxx xxx vole on June 12, 1880.  
 Weeding was no longer needed.

Prof. Darren, Mrs. Brown &  
 Mr. Cray met xxx next day xxx  
 concoct xxx plan. John Darren  
 xxx Mrs. Brown put up xxx  
 scarecrow. xxx thought xxx xxx  
 scare xxx dog. Mr. Craye put up  
 xxx fence. Problem solved.  
 Then Mrs. Brown planted lotus,  
 carnations, daffodils, xxx roses.

Next we run Map Analysis on both texts using AutoMap's default Analysis Settings

**These are the Map and Statistics  
 outputs for the first text (direct  
 adjacency):**

**These are the Map and Statistics  
 outputs for the second text (rhetorical  
 adjacency):**

Map:

5	mrs
2	mr
2	darren
2	brown
1	you
1	worked
1	with
1	week
1	weeds
1	watered
1	was
1	too
1	thought
1	then
1	take
1	stopped
1	scary
1	scarecrow
1	said
1	roses
1	prof
1	pm
1	pleased
1	plan
1	painted
1	over
1	on
1	now
1	no
1	next
1	mr
1	met
1	lotus
1	longer
1	june
1	john
1	how
1	happy
1	had
1	good
1	garden

Map:

2	mr
2	brown
1	you
1	worked
1	weeds
1	watered
1	was
1	too
1	then
1	take
1	stopped
1	said
1	prof
1	pleased
1	plan
1	over
1	on
1	now
1	no
1	next
1	mr
1	met
1	lotus
1	longer
1	june
1	john
1	happy
1	had
1	good
1	garden
1	following
1	fence
1	dog
1	discussed
1	day
1	darren
1	craye
1	cray's
1	cray
1	carnations
1	brown



	1	1880
	1	13
	1	12
Stat:	Stat:	
# of concepts analyzed:	# of concepts analyzed:	
unique: 62	unique: 60	
total: 88	total: 84	
# of concepts in statements:	# of concepts in statements:	
unique: 54	unique: 55	
total: 60	total: 61	
# of isolated concepts:	# of isolated concepts:	
unique: 12	unique: 6	
total: 13	total: 7	
# of statements:	# of statements:	
unique: 54	unique: 55	
total: 60	total: 61	
Density (based on Statements):	Density (based on Statements):	
unique: 0.87	unique: 0.92	
total: 0.97	total: 1.02	

**Note: For more information about the impact of coding choices on map analysis results you might have a look at our publications (<http://www.casos.cs.cmu.edu/projects/automap/publications.html>).**

## Content Analysis

### *From content analysis to semantic networks*

#### *I. Open AutoMap*

##### **On Empire:**

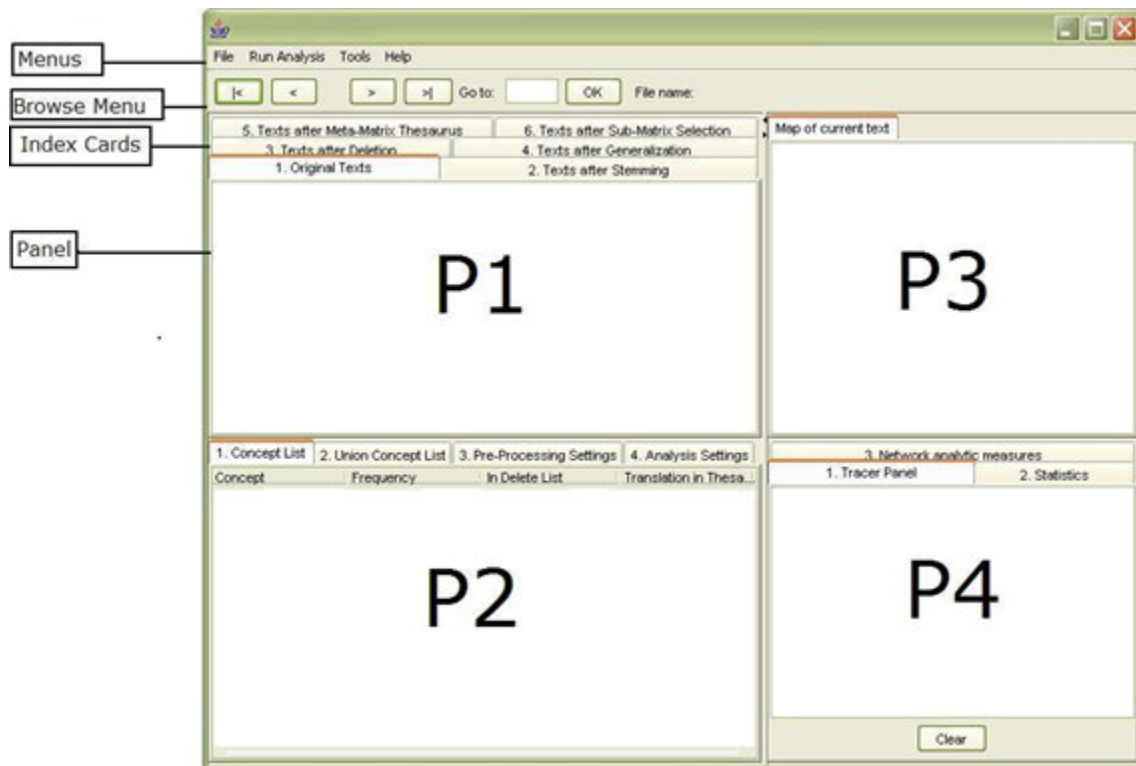
- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run Automap icon. The Automap Graphical User Interface (GUI) pops up.

From anywhere else: (requires a web connection)

<http://www.casos.cs.cmu.edu/projects/automap/software.html>

## ***II. Overview of the AutoMap Graphical User Interface (GUI)***

Below is a screen capture of the AutoMap GUI:



The AutoMap GUI is divided into four primary quadrants, or panels: P1 (top left), P2 (bottom left), P3 (top right), and P4 (bottom right).

A drop-down menu bar at the top of the window provides access to analysis tools and utilities. The Browse Menu allows you to quickly navigate between texts you have loaded into Automap.

The P2 panel can be edited, but the other panels (P1, P3, P4) cannot. Information displayed on each panel is always related to the information displayed in the other three panels. The Browse Menu relates to all four panels at the same time.

## ***III. Load multiple text documents into AutoMap***

In order to analyze multiple texts simultaneously, you will need to store the texts together in one folder.

Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextInputFiles

- In the AutoMap menu bar, select File, then Open Multiple Files. A folder chooser window will pop up.

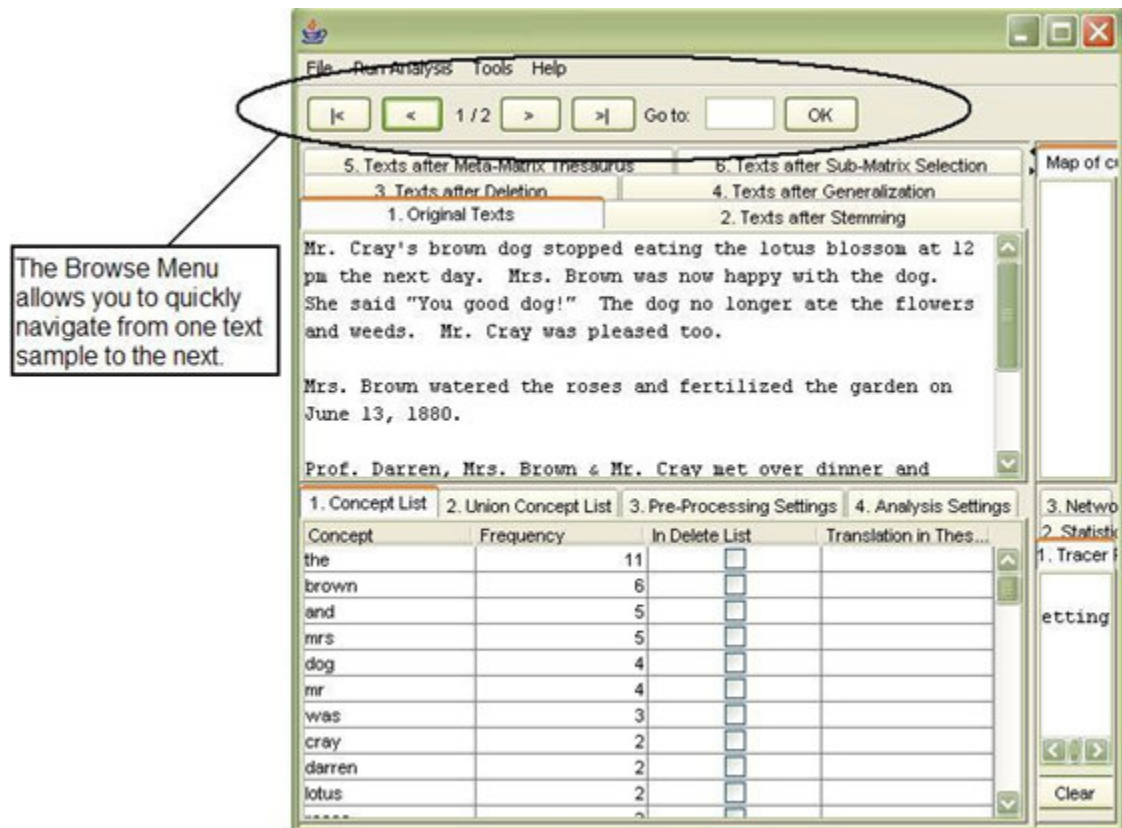
- Navigate to the folder containing the texts you wish to analyze, and single-click on it.

Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextInputFiles

- Click the Open button in the bottom right corner.

The first text will be displayed in panel P1 under the tab titled 1.Original Texts. You can use the Browse Menu to browse through the texts.

Below is a screen capture of the first text display:



#### ***IV. Carry out a simple content analysis***

A simple content analysis determines the frequencies of all words in a text document.

##### **Examine the Concept List:**

The Concept List considers concepts in each text file individually. It is created automatically when a text or set of texts is loaded into Automap, and is displayed in P2 under the tab titled 1. Concept List.

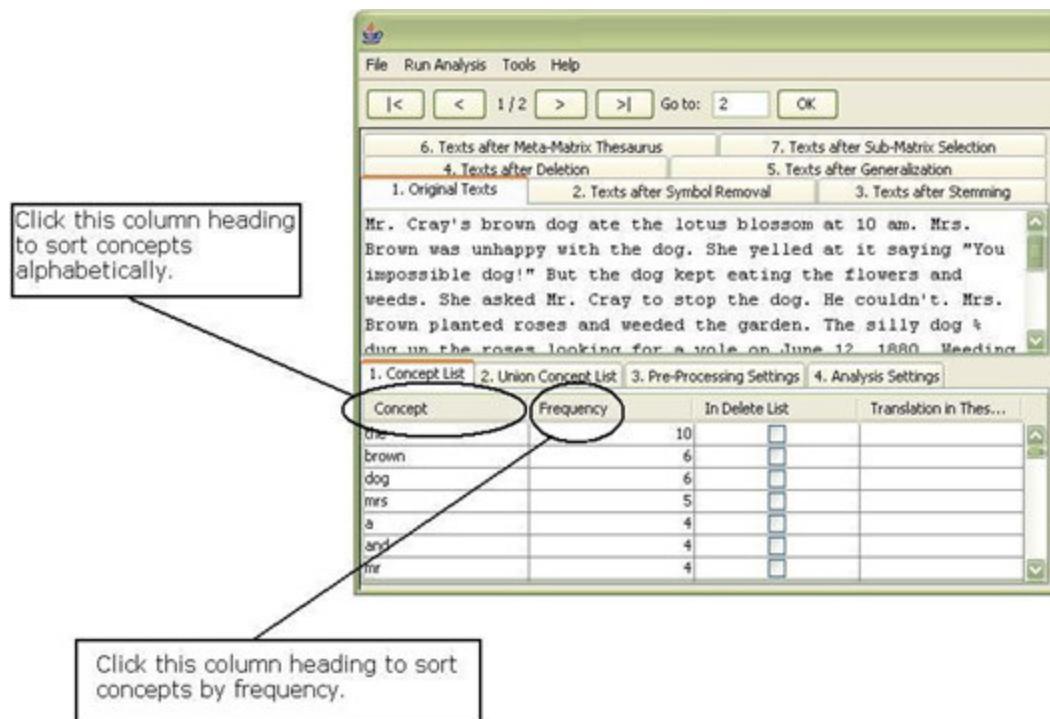
The Concept List tells you several important facts about each text in your set:

- The number of concepts found in the text displayed in P1 under the tab titled 1.Original Text.
- Related frequencies of those concepts
- Cumulated unique concepts and total concepts contained in the data set

**Note: The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts.**

By default, the Concept List is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To resort the list by frequency, click on the second-column header Frequency.

Below is a screen capture showing part of the Concept List, sorted by frequency, for the first document in our loaded folder of texts (Text1-in-user's guide.txt):



### Create and refresh the Union Concept List:

The Union Concept List considers concepts across all texts in a set (unlike the Concept List, which considers each text in the set individually). It is located in P2 under the tab titled 2.Union Concept List.

The Union Concept List tells you several important facts about your text set:

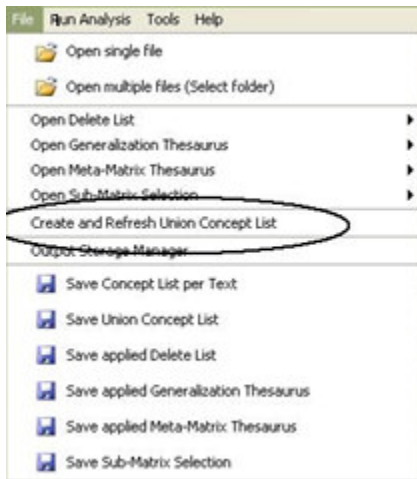
- Concepts contained in all loaded text sets

- Related, cumulative frequencies of concepts in all text sets
- Cumulative unique and total concepts.

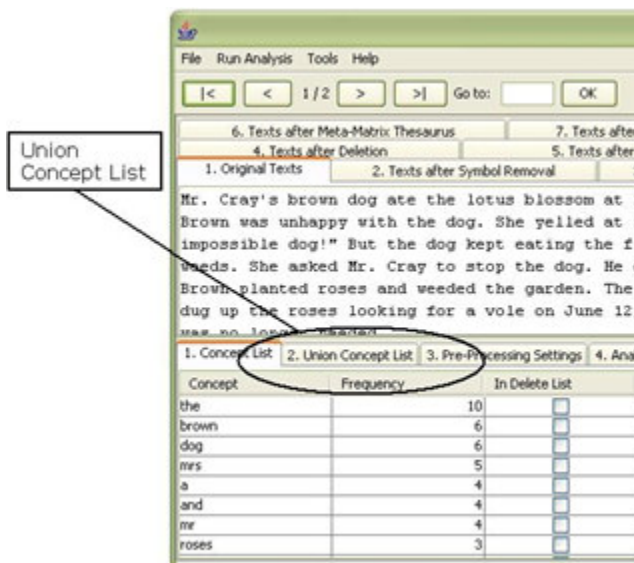
Before viewing your Union Concept List, you must refresh it. In the Menu Bar, choose File, then Refresh Union Concept List.

The Union Concept List can be refreshed after each step of pre-processing to visualize the impact of pre-processing operations on the union of concepts.

Below is a screen capture showing where to find the Refresh Union Concept List command:



After refreshing your Union Concept List, you can view it under the 2. Union Concept List tab in P2. Below is a screen capture showing where this tab is located in the GUI:



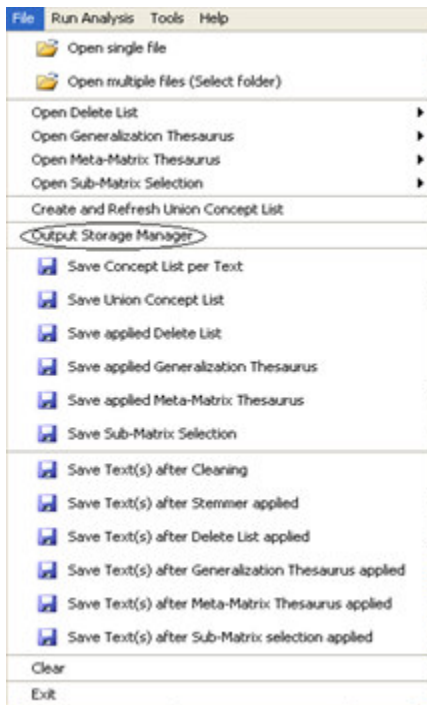
By default, the list is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To re-sort the list by Frequency, click on the second-header column Frequency.

**Note: The number of unique concepts considers each concept only once. The number of total concepts also considers repetitions of concepts.**

## ***V. Save your Union Concept List as a csv file***

To specify a location for your Union Concept List file, choose File in the Menu Bar, then Output Storage Manager.

Below is a screen capture showing where to find this in the GUI:



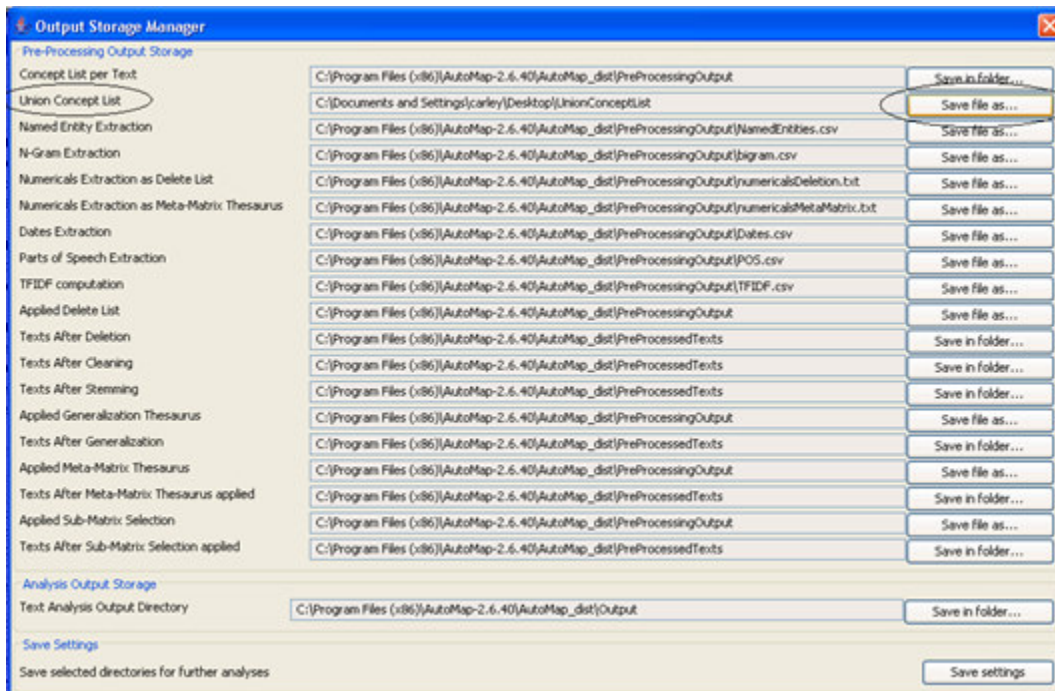
A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

The Union Concept List appears as the second item in the Output Storage Manager. Click the Save file as... button to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want.

Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\UnionConceptList

Below is a screen capture showing where to find the correct Save file as... button in the Output Storage Manager:





Click the Save Settings button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the red X at the top right corner.

In the Automap Menu Bar, choose File, then Save Union Concept List.

## ***VI. View your csv file in Microsoft Excel***

Navigate to the location of your new csv file and double-click on its icon.

Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\UnionConceptList

A box titled Open With pops up. Click on Microsoft Excel, then click OK.

**Note: You must have MS Excel installed on your computer in order to view your csv file this way.**

## ***VII. Save and view your results as a binary csv file***

Follow Steps 3 and 4.

## ***VIII. Carry out a simple Semantic Network Analysis***

A simple network analysis involves no thesauri and uses all words.

A Semantic Network Analysis can be run on original input texts or on texts that have been pre-processed with a Delete List and/or a Generalization Thesaurus.

### Before running the analysis:

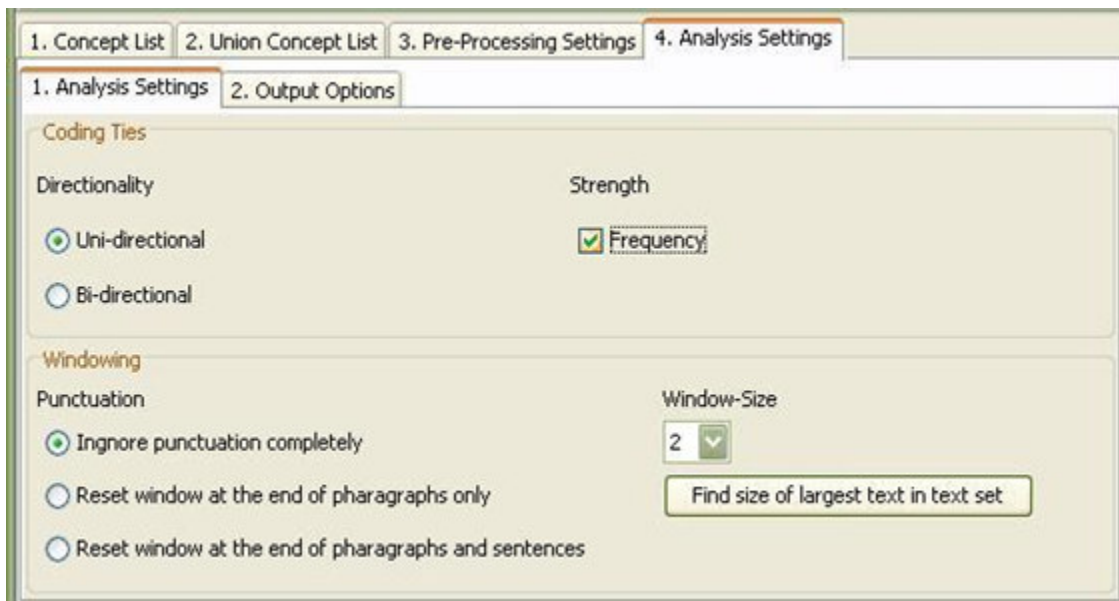
Make sure you have completed the following steps:

- Pre-process the texts if you wish to
- Specify the analysis settings (otherwise the default settings will be applied)
- Specify additional output options if you wish to (if not, the default outputs will be generated)

### About the analysis settings:

Use the Analysis Settings index card to specify the analysis settings. Your settings will automatically be applied to the analysis (you do not need to confirm them).

Below is a screen capture showing where to find the Analysis Settings index card in P2:



The following chart lists the possible analysis settings:

<b>Coding Ties</b> Specify the way statements are counted.	<b>Directionality</b> Select one of the following two possibilities by checking the button.	Uni-Directional (When coding a tie, only 1st->2nd concept should be noted)
	<b>Strength</b> Strength will be printed out by default. To not	Bi-Directional (When coding a tie, both 1st <-> 2nd concept shall be noted)
		<b>Frequency</b> (The cumulative frequency of every existing statement.)



print out Strength  
uncheck the item.

**Item not checked:**

Existence of frequency will be  
printed out (binary result).

**Ignore punctuation**

**completely** (Statements will  
be placed between all  
concepts.)

**Reset window at end of  
paragraphs only**

(Statements will be placed  
only within every single  
paragraph.)

**Reset window at end of  
paragraphs and sentences**

(Statements will be placed  
only within every single  
sentence.)

Window size between **2 and  
100**.

The Window Size defines how  
distant concepts can be and  
still have a relation ship. Only  
concepts in same window can  
form statements.

**Windowing**

Windowing is a method that  
codes the text as a map by  
placing relationships  
between pairs of Concepts  
that occur within a window.

**Punctuation**

Select one option by  
checking the radio  
button.

**Window Size**

Select one window size  
by using the number  
chooser.

If you do not want to change any of the suggested options, the analysis will be done with  
a set of standard (default) settings.

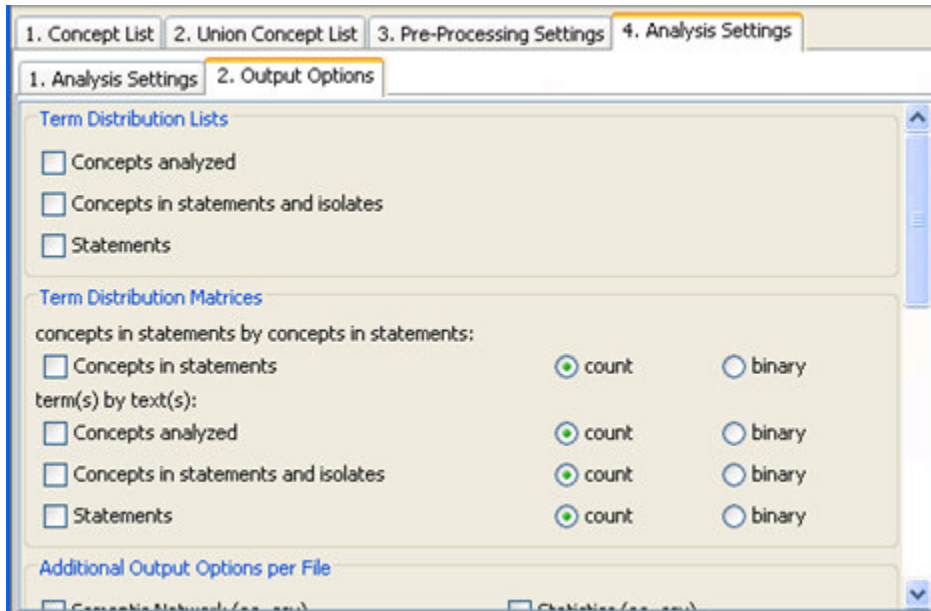
Below is a screen capture showing the standard settings:

The screenshot shows the 'Analysis Settings' tab in the AutoMap software. It contains two main sections: 'Coding Ties' and 'Windowing'. In the 'Coding Ties' section, under 'Directionality', the 'Uni-directional' radio button is selected, and under 'Strength', the 'Frequency' checkbox is unchecked. In the 'Windowing' section, under 'Punctuation', the 'Ignore punctuation completely' radio button is selected. Under 'Window-Size', the value '2' is selected in a dropdown menu, and there is a button labeled 'Find size of largest text in text set'.

**About the output options:**

The map and the statistic output generated by AutoMap are displayed in P3 and P4, respectively. In addition, AutoMap offers further output options that can be chosen in the 10.Output Options index card in P2. All additional outputs are only generated after analyses are run.

Below is a screen capture showing the 10.Output Options index card in P2:



For all types of multiple analysis, term distribution lists and matrices can be chosen as output options in the upper two fields of the 10.Output Options index card in P2.

### Points to consider:

- No list or matrix is generated by default.
- Lists or matrices are only generated if the user checks the item he or she wants and runs an analysis (of any type).
- If pre-processing was performed, the list relates to the stage of pre-processing that was used for the analysis. If several pre-processing techniques were applied, analysis will always be run on the highest stage of pre-processing.
- The requested lists and matrices are automatically saved in a folder called Term Distribution Lists and Matrices under the root directory of AutoMap. This folder is overwritten with every new analysis you run. If you want to save the results of a current Term Distribution Lists and Matrices folder, simply rename that folder.
- Two output lists are generated for each Term Distribution List checked.
- One output matrix is generated for each Term Distribution Matrix checked.

The following chart lists the types and content of Term Distribution Lists and Matrices:

Output Type		Name of output	Content of output
Term Distribution List	Concepts analyzed	List of concept analyzed.csv	Concept, Text, Frequency
		Statistics of concepts analyzed.csv	Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts
	Concepts in statements and isolates	List of concept in statements.csv	Concept, Text, Frequency
		Statistics of concept in statements.csv	Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts
		List of isolates in statements.csv	Concept, Text, Frequency
		Statistics of isolates.csv	Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts
	Statements	List of statements.csv	Statement, Text, Frequency
		Statistics of statements.csv	Statement, Cumulated sum across text set, Number of text statement occurs in, Percentage of texts statement occurs in, Texts
			Concepts in statements by concepts in statements:
	Term Distribution Matrices	Concepts in statements	Matrix of concepts in statements.csv
Term(s) by text(s):			
Concepts		Matrix of concept	Matrix of Concept (union

<b>analyzed</b>	analyzed.csv	<p>of concepts listed in first row) by texts (all text names listed in first column)</p> <p>If count was chosen, cells contain cumulated frequency of concept</p> <p>If binary was chosen, cells denote existence (1) or absence (0) of concept</p> <p>Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column)</p>
<b>Concepts in statements and isolates</b>	Matrix of concept in statements.csv	<p>If count was chosen, cells contain cumulated frequency of concept</p> <p>If binary was chosen, cells denote existence (1) or absence (0) of concept</p> <p>Matrix of Statements (union of statements listed in first row) by texts (all text names listed in first column)</p>
<b>Statements</b>	Matrix of statements.csv	<p>If count was chosen, cells contain cumulated frequency of statement</p> <p>If binary was chosen, cells denote existence (1) or absence (0) of statement</p>

### Save your upcoming analysis in DyNetML format:

Before running your Multiple Map Analysis, you must specify that your results will be in DyNetML format so that they can eventually be read into ORA.

On the 10.Output Options index card in P2, scroll down to Additional Output Formats and check the per Map box next to DyNetML for Map Analysis.

Open the Output Storage Manager. (See section V.).

You will see Text Analysis Output Directory listed under the Analysis Output Storage category. Click the Save in folder... button on the far right in that row. In the folder chooser window that pops up, navigate to the folder you want.

Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextOutputFiles

## Click Open.

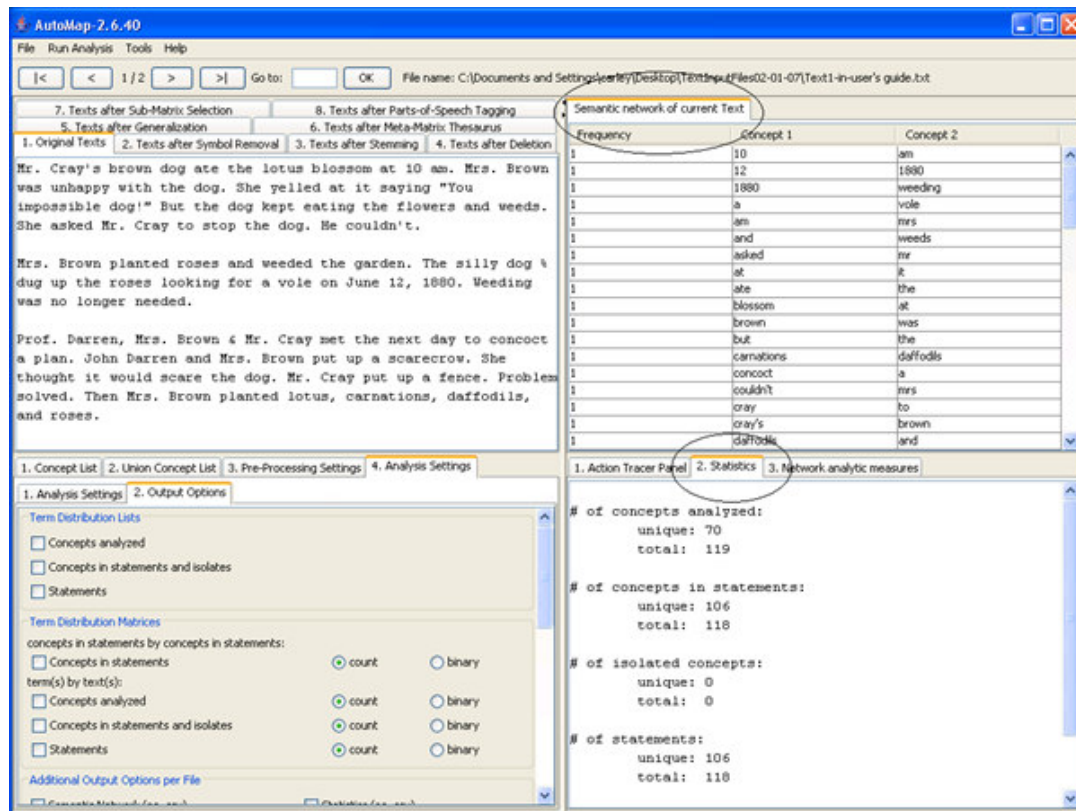
Save these settings and close the Output Storage Manager. (See section V.)

## Run the Multiple Map Analysis:

In the Menu Bar, choose Run Analysis, then Multiple Map Analysis.

The results will be displayed on the Semantic Network of Current Text index card in P3 and on the 2.Statistics index card in P4.

Below is a screen capture showing the results of Multiple Map Analysis in P3 and P4:



If you have analyzed multiple texts, you can browse through the results and see the related texts. Results of multiple analyses are automatically saved in a folder titled output under the directory where Automap is installed.

The output folders are overwritten with every new analysis you run. If you want to save the results of a current analysis folder, simply rename that folder.

## ***IX. Open ORA***

### **On Empire:**

- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run ORA icon. The ORA interface pops up.

From anywhere else: (requires a web connection)

<http://www.casos.cs.cmu.edu/projects/ora/software.html>

## ***X. Load your files into ORA***

Click the Load button on the far right. A file chooser window will pop up.

Choose Select Files. Navigate to your NyNetML files (generated in section VIII).

Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextOutputFiles

Select the files you want by holding down the key while clicking on each one. Click Open, then Finish.

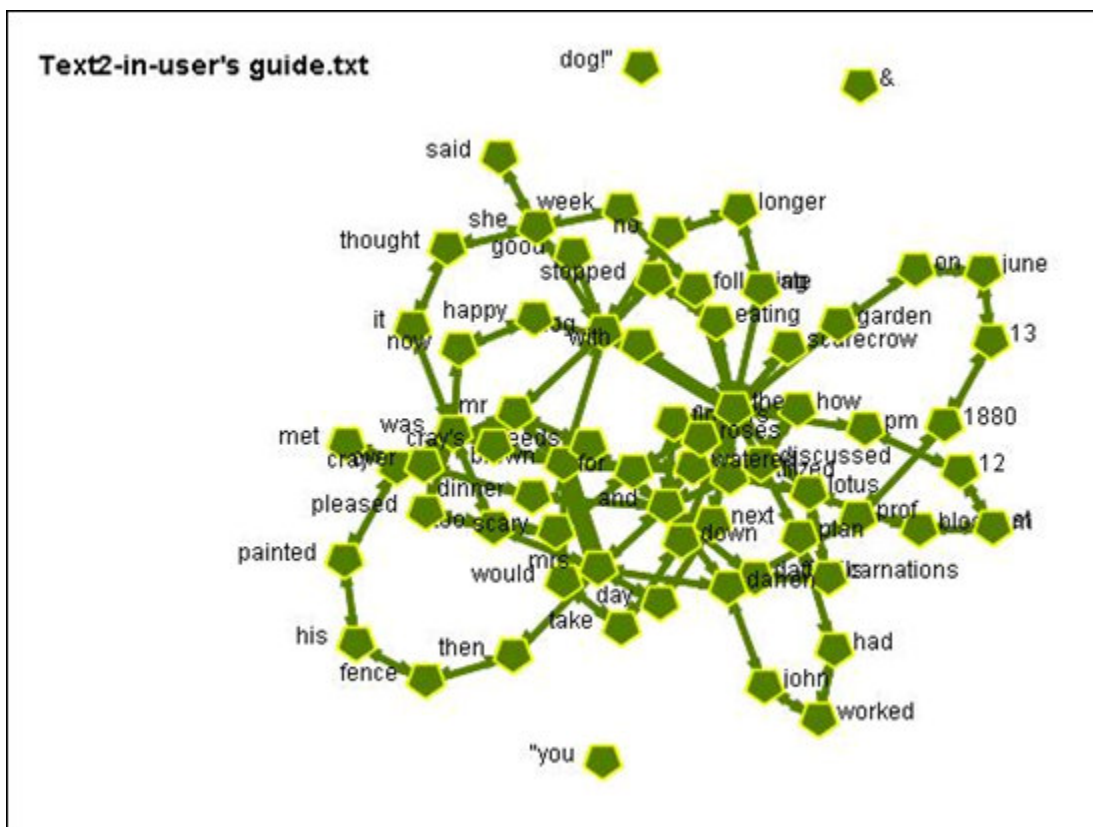
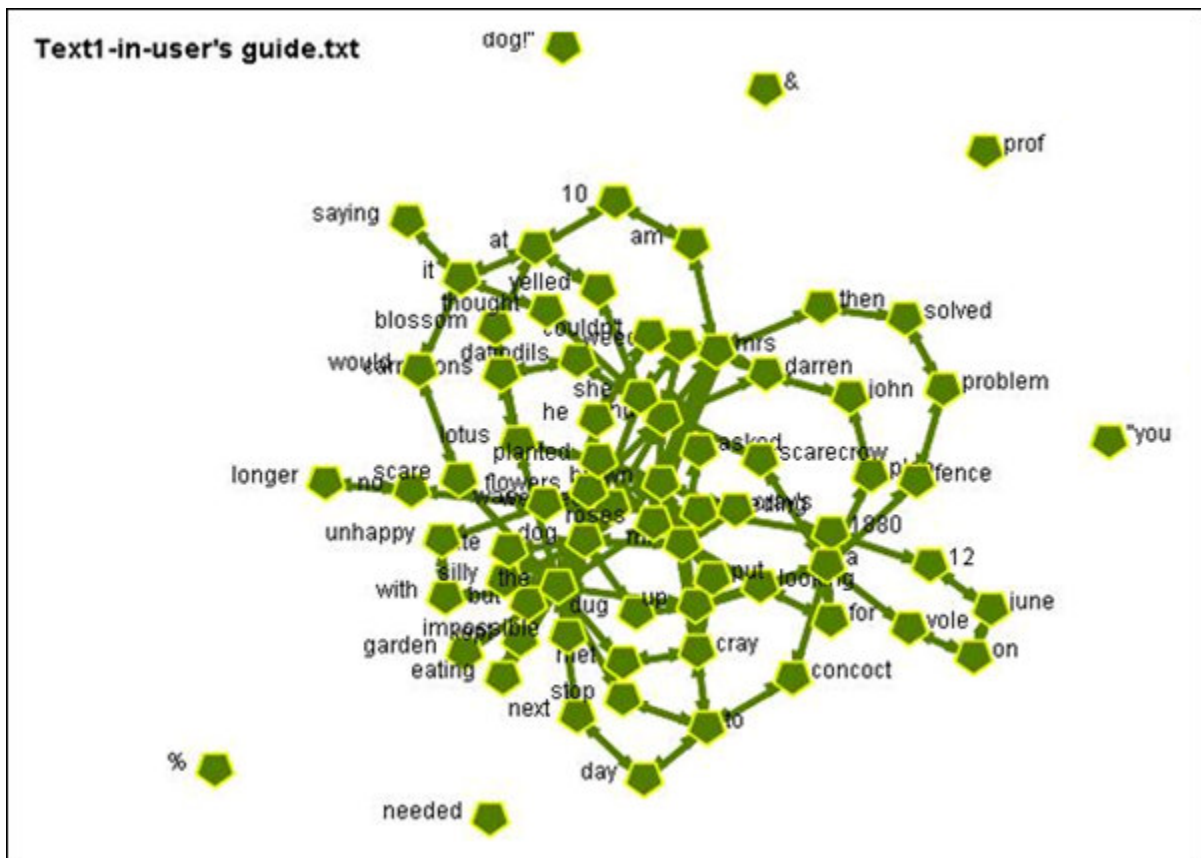
Example: Select Text1-in-user's guide.txt and Text2-in-user's guide.txt

**Note: If some file is already loaded into ORA, choose the Append As Additional Meta-Matrix option.**

## ***XI. Visualize your meta-matrices***

The left-hand portion of the ORA screen is the ORA navigator. In this navigator, select the matrix you want to visualize by clicking on it. Click the Visualize this meta-matrix button near the center of the screen to generate the visualization.

Below are the visualized meta-matrices for Text1-in-user's guide.txt (this page) and Text2-in-user's guide.txt (next page):





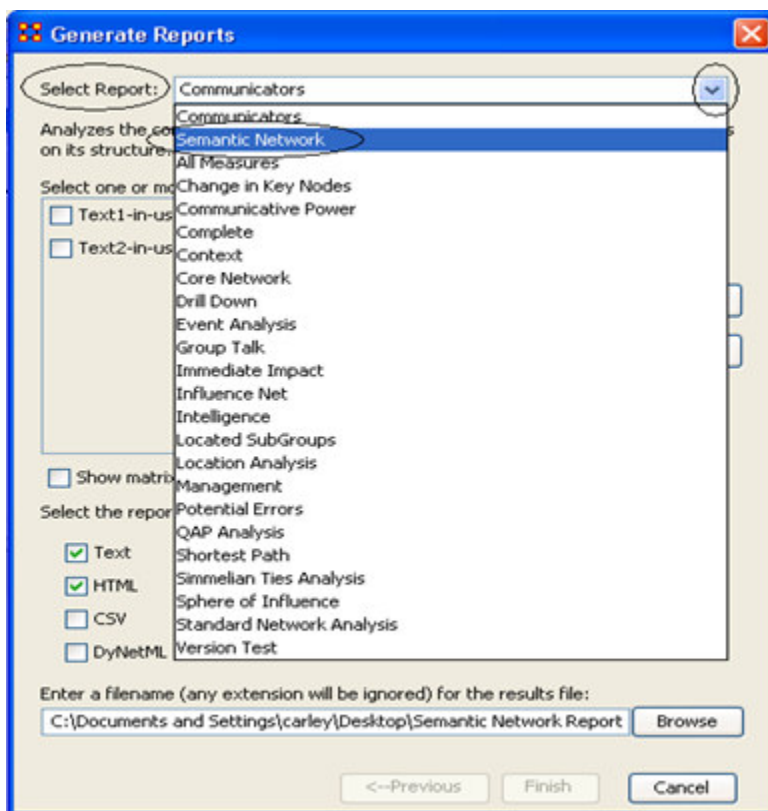
These meta-matrices also can be found in:

C:\Documents and Settings\carley\Desktop\AutomapLesson1\MetaMatrices

## ***XII. Run a Semantic Network Report***

In ORA's Menu Bar, choose Analysis, then Generate Reports. A window will pop up. You will see Select Report at the top of the window. Click the v-shaped icon to the right, and select Semantic Network from the resulting drop-down menu.

Below is a screen capture showing where to find the Semantic Network Report in the Generate Reports pop-up window:



Check the two boxes for Text1-in-user's guide.txt and Text2-in-user's guide.txt inside the field titled Select one or more meta matrices.

To specify a filename for the results, click the Browse button near the bottom right corner of the Generate Reports pop-up window. Navigate to the desired location and type in a filename or use the default name.

Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\SemanticNetworkReport.html



Click the Next button, then the Finish button at the bottom of the Generate Reports pop-up window.

ORA will run. An HTML file will pop up displaying the results.

Return to the ORA window. A small box titled Complete has popped up. Click the OK button.

Below is the Semantic Network Report generated for Text1-in-user's guide.txt and Text2-in-user's guide.txt:

---

## SEMANTIC-NETWORK REPORT

Input data: Text1-in-user's guide.txt, Text2-in-user's guide.txt

Start time: Tue Mar 06 10:59:33 2007

This is a comparison of two semantic networks. Each node in the network is considered a Concept, and each edge a Statement connecting two concepts. Statement weights are interpreted as the number of times the statement occurred in the underlying input text.

Network	Concepts	Statements	Density
Text1-in-user's guide.txt	74	200	0.0370233
Text2-in-user's guide.txt	69	184	0.0392157
Union	98	310	0.032611
Intersection	45	8	0.0040404

### ***Symmetric difference***

The symmetric distance of network A to network B is a new network that contains the entities in A that are not in B.

Network	Concepts	Statements
Text1-in-user's guide.txt to Text2-in-user's guide.txt	29	192
Text2-in-user's guide.txt to Text1-in-user's guide.txt	24	176

### ***Saved output networks***

The summary statistics across all maps saved to: C:\Documents and Settings\jmcgille\_summary\_statistics.csv

Produced by ORA developed at CASOS - Carnegie Mellon University

# Ontological Text Coding

## 1. Meta Matrix Text Analysis

Meta Matrix Text Analysis can be run on a text or a set of texts that were pre-processed with a Meta-Matrix Thesaurus. It enables the classification and analysis of concepts in texts according to the Meta-Matrix model ontology and categories of the resulting inter and intra-related sub-matrices (Diesner & Carley, 2005), Meta-Matrix Text Analysis and the social systems represented in texts. Meta-matrix based analysis of properties of social systems by investigating the inter and intra-connections between the matrices contained in the meta-matrix (cells in Table 1) can provide insight into the complex structure of social systems.

### The Meta-Matrix Model used in AutoMap

Meta-Matrix	Agent	Knowledge	Resource	Task/Event	Organization	Location	Action	Role	Attribute
Agent	Social network	Knowledge network	Capability network	Assignment network	Membership network	Agent location network			
Knowledge		Information network	Training network	Knowledge requirement network	Organizational knowledge network	Knowledge location network			
Resource			Resource network	Resource requirement network	Organizational capability network	Resource location network			
Task/Event				Precedence network	Organizational assignment network	Task/Event location network			
Organization					Interorganizational network	Organizational location network			
Location						Proximity network			
Action									
Role									
Attribute									

Before you run Meta-Matrix analysis make sure that you have completed the following actions:

- Pre-processed the text(s) with a Meta-Matrix Thesaurus. If you also want to use a Delete List and / or a Generalization Thesaurus make sure to use these tools before you use the Meta-Matrix Thesaurus.
- Specified the analysis settings. If you do not do so, the default setting default settings will be applied.
- Specified additional output options if you wish to. If you do not select additional output options, the default outputs default outputs will be generated.

## ***Analyzing Text(s)***

If you wish to analyze a single text, click the Run Analysis menu and select Single Meta Matrix Text Analysis.

If you wish to analyze a set of texts, click the Run Analysis menu and select Multiple Meta Matrix Text Analysis.

The Results will be displayed on P3 on the Map tab and on P4 on the Stat tab respectively. If you had requested additional outputs, those will be generated and stored under the directories specified under the Section Additional Outputs.

If you have analyzed multiple texts, you can browse through the results and see the related texts.

Results of multiple analyses are automatically saved in a folder named “output” under the directory where AutoMap 2.0 is installed. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as stat\_output.xls file a that contains the stats of all texts.

The “Stat Output” folder is overwritten with every new analysis you run. So if you want to save the results of a current “Stat Output” folder just rename the folder.

### ***1.1 Example for Meta Matrix Text Analysis***

An extract from the Our Text I.txt was used as input:

Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

The text was pre-processed with the Meta-Matrix-Thesaurus. This are the resulting texts:

Setting	Resulting text
Not Selected - Thesaurus content only (default)	<p>mr . &lt; agent&gt; &lt; agent&gt; dog ate lotus &lt; event&gt; &lt; attribute&gt; am . mrs . &lt; agent&gt; was unhappy with dog . yelled saying &lt; agent resource&gt; impossible dog!" dog kept eating flowers weeds . asked mr . cray stop dog . couldn't . mrs . &lt; agent&gt; planted roses weeded garden . silly dog &lt; attribute&gt; dug up roses looking vole on june &lt; attribute&gt; , &lt; attribute&gt; . weeding was no longer needed . prof . &lt; agent&gt; , mrs . &lt; agent&gt; &lt; attribute&gt; mr . cray met next day &lt; task&gt; plan . john &lt; agent&gt; mrs . &lt; agent&gt; put up scarecrow . thought scare dog . mr . &lt; agent&gt; put up fence . problem solved . then mrs . &lt; agent&gt; planted lotus , &lt; resource&gt; , &lt; resource&gt; , roses .</p>
Thesaurus content only, Direct Adjacency (default if Thesaurus content only is chosen)	<p>. &lt; agent&gt; &lt; agent&gt; &lt; event&gt; &lt; attribute&gt; . . &lt; agent&gt; . &lt; agent resource&gt; . . . . &lt; agent&gt; . &lt; attribute&gt; &lt; attribute&gt; &lt; attribute&gt; . . . &lt; agent&gt; . &lt; agent&gt; &lt; attribute&gt; . &lt; task&gt; . &lt; agent&gt; . &lt; agent&gt; . . . &lt; agent&gt; . . . &lt; agent&gt; &lt; resource&gt; &lt; resource&gt; .</p> <p>xxx . &lt; agent&gt; &lt; agent&gt; xxx xxx xxx &lt; event&gt; &lt; attribute&gt; xxx . xxx . &lt; agent&gt; xxx xxx xxx xxx . xxx xxx &lt; agent resource&gt; xxx xxx xxx xxx xxx xxx xxx . xxx xxx . xxx xxx xxx . xxx . xxx . &lt; agent&gt; xxx xxx xxx xxx . xxx xxx &lt; attribute&gt; xxx xxx xxx xxx xxx xxx xxx &lt; attribute&gt; &lt; attribute&gt; . xxx xxx xxx xxx xxx . xxx xxx xxx . &lt; agent&gt; xxx . &lt; agent&gt; &lt; attribute&gt; xxx . xxx xxx xxx xxx &lt; task&gt; xxx . xxx &lt; agent&gt; xxx . &lt; agent&gt; xxx xxx xxx . xxx xxx xxx . xxx . &lt; agent&gt; xxx xxx . xxx xxx . xxx xxx . &lt; agent&gt; xxx xxx &lt; resource&gt; &lt; resource&gt; xxx .</p>
Thesaurus content only, Rhetorical Adjacency	

Then we run Map Analysis on both texts using AutoMap's default Analysis Settings. These are the results:

	Not Selected - Thesaurus content only (default)	Thesaurus content only, Direct Adjacency	Thesaurus content only, Rhetorical Adjacency
Map	1 10 am	1 10 am	1 blossom 10
	1 12 1880	1 12 1880	1 brown was
	1 1880 weeding	1 1880 weeding	1 carnations daffodils
	1 am mrs	1 am mrs	1 concoct plan
	1 asked mr	1 asked mr	1 couldn't mrs
	1 ate lotus	1 ate lotus	1 cray stop
	1 blossom 10	1 blossom 10	1 cray's brown
	1 brown was	1 brown was	1 craye put up
	1 carnations daffodils	1 carnations daffodils	1 daffodils roses
	1 concoct plan	1 concoct plan	1 day concoct
	1 couldn't mrs	1 couldn't mrs	1 dog yelled
	1 cray stop	1 cray stop	1 dug up
	1 cray's brown	1 cray's brown	1 eating flowers
	1 craye put up	1 craye put up	1 fence problem

1 daffodils roses  
 1 day concoct  
 1 dog yelled  
 1 dug up  
 1 eating flowers  
 1 fence problem  
 1 flowers weeds  
 1 garden silly  
 1 impossible dog  
 1 john darren  
 1 june 12  
 1 kept eating  
 1 longer needed  
 1 looking vole  
 1 lotus carnations  
 1 met next  
 1 mr craye  
 1 needed prof  
 1 next day  
 1 no longer  
 1 on june  
 1 plan john  
 1 planted roses  
 1 problem solved  
 1 prof darren  
 1 put up  
 1 put up fence  
 1 roses weeded  
 1 saying you  
 1 scare dog  
 1 scarecrow thought  
 1 silly dog  
 1 solved then  
 1 stop dog  
 1 then mrs  
 1 thought scare  
 1 unhappy with  
 1 up scarecrow  
 1 vole on  
 1 was unhappy  
 1 weeded garden  
 1 weeding was  
 1 weeds asked  
 1 with dog  
 1 yelled saying  
 1 you impossible  
 2 brown planted  
 2 darren mrs  
 2 mr cray

1 daffodils roses  
 1 day concoct  
 1 dog yelled  
 1 dug up  
 1 eating flowers  
 1 fence problem  
 1 flowers weeds  
 1 garden silly  
 1 impossible dog  
 1 john darren  
 1 june 12  
 1 kept eating  
 1 longer needed  
 1 looking vole  
 1 lotus carnations  
 1 met next  
 1 mr craye  
 1 needed prof  
 1 next day  
 1 no longer  
 1 on june  
 1 plan john  
 1 planted roses  
 1 problem solved  
 1 prof darren  
 1 put up  
 1 put up fence  
 1 roses weeded  
 1 saying you  
 1 scare dog  
 1 scarecrow thought  
 1 silly dog  
 1 solved then  
 1 stop dog  
 1 then mrs  
 1 thought scare  
 1 unhappy with  
 1 up scarecrow  
 1 vole on  
 1 was unhappy  
 1 weeded garden  
 1 weeding was  
 1 weeds asked  
 1 with dog  
 1 yelled saying  
 1 you impossible  
 2 brown planted  
 2 darren mrs  
 2 mr cray  
 5 mrs brown

1 flowers weeds  
 1 garden silly  
 1 impossible dog  
 1 john darren  
 1 june 12  
 1 kept eating  
 1 longer needed  
 1 looking vole  
 1 lotus carnations  
 1 met next  
 1 mr craye  
 1 needed prof  
 1 next day  
 1 no longer  
 1 on june  
 1 plan john  
 1 planted roses  
 1 problem solved  
 1 prof darren  
 1 put up  
 1 put up fence  
 1 roses weeded  
 1 saying you  
 1 scare dog  
 1 scarecrow thought  
 1 silly dog  
 1 solved then  
 1 stop dog  
 1 then mrs  
 1 thought scare  
 1 unhappy with  
 1 up scarecrow  
 1 vole on  
 1 was unhappy  
 1 weeded garden  
 1 weeding was  
 1 weeds asked  
 1 with dog  
 1 yelled saying  
 1 you impossible  
 2 brown planted  
 2 darren mrs  
 2 mr cray  
 5 mrs brown

Stat File:  
 # of concepts analyzed:

File:

File:

unique: 62 total: 88	# of concepts analyzed: unique: 62 total: 88	# of concepts analyzed: unique: 62 total: 88
# of concepts in statements: unique: 80 total: 87	# of concepts in statements: unique: 80 total: 87	# of concepts in statements: unique: 80 total: 87
# of isolated concepts: unique: 0 total: 0	# of isolated concepts: unique: 0 total: 0	# of isolated concepts: unique: 0 total: 0
# of statements: unique: 80 total: 87	# of statements: unique: 80 total: 87	# of statements: unique: 80 total: 87
Density (based on Statements): unique: 1.29 total: 1.4	Density (based on Statements): unique: 1.29 total: 1.4	Density (based on Statements): unique: 1.29 total: 1.4

Note: For more information about the impact of coding choices on map analysis results please visit us on the web <http://www.casos.cs.cmu.edu/projects/automap/publications.html>).

## 2. Sub Matrix Text Analysis

Sub Matrix Text Analysis distills one or several sub-networks from the meta-matrix network and retranslates the meta-matrix entities into the text-level concepts that represent these Meta-Matrix categories. This routine enables a more thorough analysis of particular cells of the meta-matrix (*Diesner & Carley, 2004c*). Sub Matrix Text Analysis can be run on a text or a set of texts that were pre-processed with a Meta-Matrix Thesaurus and from that Sub-Matrices were selected.

Before you run Sub-Matrix analysis make sure that you have completed the following actions:

- Pre-processed the text(s) with a Meta-Matrix Thesaurus. If you also want to use a Delete List and/ or a Generalization Thesaurus make sure to use these tools before you use the Meta-Matrix Thesaurus.
- Selected Sub-Matrix Categories.
- Specified the analysis settings. If you do not do so, the default setting default settings will be applied.
- Specified additional output options if you wish to. If you do not select additional output options, the default outputs default outputs will be generated.

The user is given the option to exclude entities of any applied ontology from being considered as nodes and thus forming statements, but are considered as attributes of other entities in the ontology that are forming entities and link into statements. An example would be the category “attribute”, which would represent information that is inherent to a certain entity. The concept “teacher” or “male” might be considered as attributes that relate to the entity agent.

To use ontologies (in a way ontologies not only represent entities, but also features of entities) follow these steps:

1. Use the drag and drop labeled Concepts not forming edges at the bottom of P2, tab no 10. Output Options. Click Refresh to automatically generate a list of all entities in the applied ontology.
2. Use the > and < buttons to move entities from one window to the other.
3. Apply your decision by hitting the Apply Selection button.
4. Before running Sub-Matrix Text Analysis decide whether you want to perform “Network Text Analysis” (NTA) or “Social Network Text Analysis” (SNTA). To select a type use the toggle button at the bottom of P2, 10. Output Options index card. Both types are a form of Sub-Matrix Text Analysis, but differ in what they measure:
  1. NTA: If one measures textual network, then entities of the applied ontology that should not be considered as nodes, but as inherent information of nodes, fill structural position and semantic function in a text. Thus they contribute to a texts' density, and therefore should be taken into consideration as statements for maps and stats. The resulting DyNetML file will contain nodes and inherent information on nodes (if any contained in the window that an entity and an attribute co-occur).
  2. SNTA: If one measures a social network that is represented in or extracted from a text, then links from entities of the applied ontology that should not be considered as nodes, but as inherent information of nodes, to actual other nodes are not to be represented in the maps and stats, since this would over fit the social network. The resulting DyNetML file will NOT inherent information on nodes.  
Whether to use NTA or SNTA is a “text-philosophical question” that's answer depends upon what the user wants to measure - a textual network or a social network.
5. Run Sub-Matrix Text Analysis.

If you wish to analyze a single text, click the Run Analysis menu and select Single Meta-Matrix analysis.

If you wish to analyze a set of texts, click the Run Analysis menu and select Multiple Meta-Matrix Analysis.

The Results will be displayed on P3 on the Map index card and on P4 on the Stat index card.

If you had requested additional outputs, those will be generated and stored under the directories specified under the Section Additional Outputs.

If you have analyzed multiple texts, you can browse through the results and see the related texts.

Results of multiple analysis are automatically saved in a folder named “output” under the directory where AutoMap 2.0 is installed. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as stat\_output.xls file a that contains the stats of all texts.

The “Stat Output” folder is overwritten with every new analysis you run. So if you want to save the results of a current “Stat Output” folder just rename the folder.

## ***2.1 Example for Sub Matrix Text Analysis***

This example is based on the example for Sub Matrix Selection.

Our Text I.txt was used as input:

Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

The text was pre-processed with the Meta-Matrix-Thesaurus. Then we selected the full meta-matrix. These are the resulting texts:

### **Setting**

Select:  
Knowledge/Agent/Resource

### **Resulting text**

. dog lotus.. dog. dog flowers.. dog... roses. dog roses,..

.,..... dog..... lotus, carnations, daffodils, roses.

xxx. xxx xxx dog xxx xxx lotus xxx xxx xxx xxx. xxx.

xxx xxx xxx xxx xxx dog. xxx xxx xxx xxx xxx xxx xxx

xxx xxx xxx dog xxx xxx xxx flowers xxx xxx. xxx xxx

xxx. xxx xxx xxx xxx dog. xxx xxx. xxx. xxx xxx roses

Select: xxx xxx xxx xxx. xxx xxx dog xxx xxx xxx xxx roses xxx

Knowledge/Agent/Resource xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx

xxx xxx. xxx, xxx. xxx xxx xxx. xxx xxx xxx xxx xxx

xxx xxx xxx xxx. xxx xxx xxx xxx. xxx xxx xxx xxx xxx.

xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx

xxx. xxx xxx. xxx xxx lotus, carnations, daffodils, xxx



roses. xxx

Then we run Map Analysis on both texts using AutoMap's default Analysis Settings.  
These are the results:

1 lotus carnations	1 12 pm
1 met the	1 13 1880
1 mr craye	1 1880 prof
1 neededprof darren	1 and weeds
1 next day	1 at 12
1 no longer	1 ate the
1 on june	1 blossom at
1 plan john	1 brown would
1 planted roses	1 carnations daffodils
1 problem solved	1 cray was
1 put up	1 cray's brown
1 put up a	1 craye painted
1 roses looking	1 daffodils and
1 saying you	1 darren mrs
1 scare the	1 day mrs
1 scarecrow she	1 dinner and
1 she yelled	1 discussed how
1 silly dog	1 dog the
Map 1 solved then	1 down the
1 stop the	1 eating the
1 the silly	1 fence then
1 then mrs	1 fertilized the
1 thought it	1 flowers and
1 to stop	1 following week
1 unhappy with	1 for the
1 up the	1 garden on
1 vole on	1 good dog
1 was unhappy	1 had worked
1 weeded the	1 happy with
1 weeding was	1 his fence
1 weeds she	1 how the
1 with the	1 it was
1 would scare	1 john darren
1 yelled at	1 june 13
1 you impossible	1 longer ate
2 brown planted	1 lotus carnations
2 mr cray	1 met over

4 the dog  
5 mrs brown

1 mr craye  
1 next day  
1 no longer  
1 now happy  
1 on june  
1 over dinner  
1 painted his  
1 plan had  
1 pleased too  
1 pm the  
1 prof darren  
1 roses and  
1 said you  
1 scarecrow the  
1 scary for  
1 she thought  
1 stopped eating  
1 take down  
1 the scarecrow  
1 then mrs  
1 thought it  
1 too scary  
1 was too  
1 watered the  
1 weeds mr  
1 week she  
1 with the  
1 worked john  
1 would take  
1 you good  
2 brown watered  
2 mr cray  
3 the dog  
5 mrs brown

# of concepts analyzed:  
unique: 72  
total: 118

# of concepts analyzed:  
unique: 68  
total: 107

# of concepts in statements:  
unique: 108  
total: 117

# of concepts in statements:  
unique: 98  
total: 106

Stat # of isolated concepts:  
unique: 0  
total: 0

# of isolated concepts:  
unique: 0  
total: 0

# of statements:  
unique: 108  
total: 117

# of statements:  
unique: 98  
total: 106

Density (based on Statements):	Density (based on Statements):
unique: 1.5	unique: 1.44
total: 1.63	total: 1.56

## Outputs

Outputs for Map Analysis, Meta Matrix Text Analysis and Sub Matrix Text Analysis are displayed on P3 on the Semantic Network index card and on P4 on the Stat index card.

If you have analyzed multiple texts, you can browse through the results and see the related texts.

Results of multiple analyses are automatically saved in a folder named “output” under the directory where AutoMap 2.0 is installed. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as a stat\_output.xls file that contains the stats of all texts. Additional Outputs will be generated if requested by the user.

### **1. Semantic Network**

After running analysis, the semantic network will be displayed on P3 on the Semantic Network Index card.

The semantic network contains one coded statement per line.

If the Frequency item was checked on the Analysis Settings index card the first column of the semantic network indicates the frequency of every displayed statement.

Each semantic network generated is automatically saved in a folder named “Stat Output” under the directory where AutoMap 1.2 is installed. This output folder contains a semantic network file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as a stat\_output.xls file that contains the stats of all texts.

#### **1.1 Example for Semantic Network Output**

For examples for semantic network files, see the examples for analysis, the Semantic Network of current text index cards.

Semantic networks are output as .CSV (.map in previous versions).

### **2. Statistics**

After running the analysis, the Stat file will be displayed on P4 on the Statistics Index Card.

Each stat file generated is automatically saved in a folder named “Stat Output” under the directory where AutoMap 1.2 is installed. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as stat\_output.xls file a that contains the stats of all texts.

Entries in the stat output and explanation:

Entry	Entry	Explanation
File:		Name of the analyzed text file.
	unique:	Unique concepts are those that appear only once in a text; the number of total concepts includes those that appear more than once in a given text.
# of concepts analyzed:	total:	All concepts are considered that occurred in the texts that were analyzed.
	unique:	Only concepts are considered that linked into statements.
# of concepts in statements:	total:	
	unique:	Only concepts are considered that did not link into statements.
# of isolated concepts:	total:	
	unique:	Unique statements are those that appear only once in a text; the number of total statements includes those that appear more than once in a given text.
# of statements:	total:	
	unique:	Unique density is the density of the resulting network based on unique statements, total density respectively is the density of the resulting network based on the total number of statements.
density (based on statements)	total:	
	Punctuation:	The Punctuation option chosen by the user.
Analysis Settings	Window Size:	The Window Size chosen by the user.
	Directionality:	The Directionality option chosen by the user.

## ***2.1 Example for Statistic Output***

For examples of stat files, see the examples for analysis, the Stat of current text index cards.

## ***3. Additional Outputs***

The map and the statistic output generated by AutoMap are displayed on P3 and P4, respectively.

For all types of multiple analysis a lot more outputs can be generated on demand.

## **Output Options**

The map and the statistic output generated by AutoMap are displayed on P3 and P4, respectively. In addition to that AutoMap offers further output options that can be chosen on P2, 10. Output Options index card. All additional outputs are only generated after analyses were run.

Pre-processing and Analysis Settings

1. Concept List | 2. Union Concept List | 3. Utilities | 4. Stemming | 5. Delete List | 6. Generalization Thesaurus  
7. Meta-Matrix Thesaurus | 8. Sub-Matrix Selection | 9. Analysis Settings | 10. Output Options

**Term Distribution Lists**

☐ Concepts analyzed

☐ Concepts in statements and isolates

☐ Statements

**Term Distribution Matrices**

concepts in statements by concepts in statements:

☐ Concepts in statements ☒ count ☐ binary

term(s) by text(s):

☐ Concepts analyzed ☒ count ☐ binary

☐ Concepts in statements and isolates ☒ count ☐ binary

☐ Statements ☒ count ☐ binary

**Non-identified Concepts**

☐ Save list of Non-identified Concepts

**Additional Output Formats**

Ucinet (DL) ☐ Maps ☐ Term Distribution Matrices

DyNetML for Map Analysis ☐ per Map

DyNetML for Sub-Matrix Analysis ☐ per Text ☐ per TextSet

**Network analytic measures**

☐ Measures

**Concepts not forming edges**

☒ Network Text Analysis ☐ Social Network Text Analysis

## 1. Term Distribution Lists and Matrices

Term distribution list and matrices as output options for all types of multiple analysis can be chosen in the upper two fields of P2, tab no. 10. Output Options.

### Points To Consider

- No list or matrix will be generated by default.
  - All list or matrix are only generated if the user checks a list he or she wants to be generated and runs an analysis of any type.
  - If pre-processing was performed, the list relates to the stage of pre-processing that was used for the analysis. If several pre-processing techniques were applied, analysis will always be run on the highest stage of pre-processing.
  - The requested lists and matrices will be automatically saved in a folder called Term Distribution Lists and Matrices under the root directory of AutoMap.
  - The Term Distribution Lists and Matrices folder is overwritten with every new analysis you run. So if you want to save the results of a current Term Distribution Lists and Matrices folder just rename the folder.
- For each Term Distribution List checked two output lists are generated.

- For each Term Distribution Matrix checked one output matrix is generated.  
See the output examples for the content of each list.

## 1.1 Types and content of Term Distribution Lists and Matrices

Output Type	Name of output	Content of output
Term Distribution List	<b>Concepts analyzed</b>	List of concept analyzed.csv Concept, Text, Frequency  Statistics of concepts analyzed.csv Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts  List of concept in statements.csv Concept, Text, Frequency  Statistics of concept in statements.csv Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts  List of isolates in statements.csv Concept, Text, Frequency  Statistics of isolates.csv Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts  List of statements.csv Statement, Text, Frequency  <b>Statements</b> Statistics of statements.csv Statement, Cumulated sum across text set, Number of text statement occurs in, Percentage of texts statement occurs in, Texts
	<b>Concepts in statements and isolates</b>	
	<b>Concepts in statements</b>	
	<b>Concepts in statements by concepts in statements:</b>	
	<b>Concepts in statements</b>	
	<b>Concepts in statements</b>	
	<b>Concepts in statements</b>	
	<b>Concepts in statements</b>	
	<b>Concepts in statements</b>	
	<b>Concepts in statements</b>	
	<b>Concepts in statements</b>	
Term Distribution Matrices	<b>Concepts in statements</b>	Matrix of concepts in statements.csv Matrix of Concept that were linked into statements ( first row) by Concept that were linked into statements (first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
	<b>Concepts analyzed</b>	Matrix of concept analyzed.csv Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
	<b>Concepts in</b>	Matrix of concept in Matrix of Concept (union of concepts listed in first row) by texts (all text

<b>statements and isolates</b>	statements.csv	names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
<b>Statements</b>	Matrix of statements.csv	Matrix of Statements (union of statements listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of statement If binary was chosen, cells denote existence (1) or absence (0) of statement

## 1.2 Example for Term Distribution List and Matrices

Extracts from the Denmark.txt and USA.txt files were used as input:

Our Text 1.txt: Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

Our Text II.txt: Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray was pleased too. Mrs. Brown watered the roses and fertilized the garden on June 13, 1880. Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed how the plan had worked. John Darren and Mrs. Brown would take down the scarecrow the following week. She thought it was too scary for the dog. Mr. Craye painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.

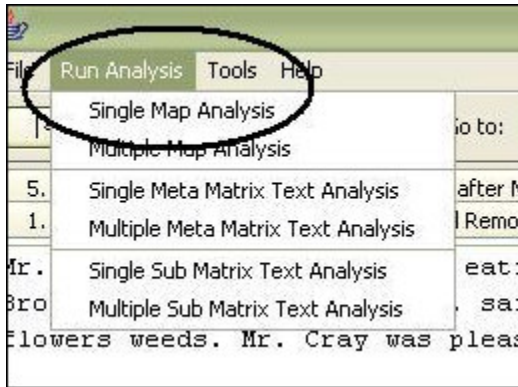
Then AutoMap's Extensive Delete List, Direct Adjacency was applied to both texts. The Delete List was extended by further non-content bearing words that appeared in the sample texts (a an and as at awhile but for from happening he her her hers him his i in into it its me mine my nor of or our she so that the their theirs them they to us was we were what who whoever whom whomever will would you your yours yourself). Below is the resulting texts:

Our Text 1.txt: Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.



Our Text II.txt: Mr. Cray's brown dog stopped eating lotus blossom 12 pm next day. Mrs. Brown was now happy with dog. said "You good dog!" dog no longer ate flowers weeds. Mr. Cray was pleased too. Mrs. Brown watered roses fertilized garden on June 13, 1880. Prof. Darren, Mrs. Brown & Mr. Cray met over dinner discussed how plan had worked. John Darren Mrs. Brown take down scarecrow following week. thought was too scary dog. Mr. Cray painted fence. Then Mrs. Brown watered lotus, carnations, daffodils, roses.

Then semantic network analysis was run on both texts using AutoMap's default settings. The black ellipse in the screen shot below shows how to run a Single Map Analysis (Semantic Network Analysis):



All outputs provided on the Output Options panel were checked. Again, the default settings were used. All output lists are saved in this folder (Term\_Distribution\_Lists\_and\_Matrices.zip).

## ***2. Save Non-Identified concepts***

Purpose: Save a list of all concepts that are remaining in the pre-processed texts and that are not:

- Denoted in a delete list
- Denoted in any of the thesauri

To create a list of these words, go to on >span>P2, tab no. 4 - Analysis Settings, tab no. 2 - Output Options index card, select the Non-Identified Concepts field and check Save list of Non-Identified Concepts. Run any type of Analysis. The list will be stored under the root directory of AutoMap as Non\_identified\_concepts.csv. The black ellipse in the screen shot below shows where to find this option.

The screenshot shows the 'Analysis Settings' tab in the AutoMap software. The 'Non-Identified Concepts' section is highlighted with a black oval, indicating that the checkbox 'Save list of Non-identified Concepts' is checked. Other sections visible include 'Additional Output Options per File' with checkboxes for 'Semantic Network (as .csv)' and 'Statistics (as .csv)'; 'Additional Output Formats' with checkboxes for 'UciNet (DL)', 'Maps', 'Term Distribution Matrices', 'DyNetML for Map Analysis', 'per Map', 'DyNetML for Sub-Matrix Analysis', 'per Text', and 'per TextSet'; 'Network analytic measures' with a checkbox for 'Measures'; and 'Sub Matrix Analysis Options'.

### 3. Additional Output Formats

Two additional data formats are offered:

- DL for e.g. UCINET (Borgatti, Everett & Freeman, 2002; for more information about UCINET see <http://www.analytictech.com/ucinet.htm>).
- DyNetML (Tsvetovat, Reminga & Carley, 2004; for more information about DyNetML see <http://www.casos.cs.cmu.edu/projects/dynetml/>).

**Note: You can also use the Network Converter to convert data.**

#### 3.1 DL:

If maps and/or term distribution matrices are generated, these files can be additionally stored in the UCINET DL format.

Some points to consider when storing in UCINET DL Format:

- In order to do this check "Maps" and/ or "Term distribution matrices" items on the "Additional Output Formats" field on P2, tab no. 10 Output Options index card.
- The maps are stored with the extension .dl in the "DL\_UciNet\_Format" folder in the root directory of AutoMap.
- The term distribution matrices are stored with the extension .dl in the "DL\_UciNet\_Format" folder in the root directory of AutoMap.
- If you wish to perform UCINET analysis on your AutoMap results, import the data into UCINET. If you wish to visualize your AutoMap results, import the data into UCINET and use the visualization tools provided through UCINET.

**Note: Since networks extracted with AutoMap are directed, matrices representing these networks can be rectangular. If a DL file needs to be generated from a rectangular matrix AutoMap by default squares this matrix before converting it into DL format.**

### **3.2 DyNetML:**

Results of Map Analysis and Sub-Matrix Text Analysis can be output in DyNetML format.

Purpose: Generate DyNetML representation of maps (mental models) generated with Map Analysis.

How to: Check the "per Map" checkbox on P2, 10. Output Options index card, Additional Output Formats, DyNetML for Map Analysis.

Output: The resulting DyNetML files will be stored as NameOfText.xml in the xml folder under the root directory of AutoMap.

After Map Analysis:

By default, all entities in the applied ontology are considered as entities, and all statements between entities in the applied ontology (either in anterior or posterior or both positions) are represented as edges in DyNetML. The user is given the option to exclude entities of any applied ontology from being considered as nodes and thus forming statements, but are considered as attributes of other entities in the ontology that are forming entities and link into statements. An example would be the category "attribute", which represent information that is inherent to a certain entity. The concept "teacher" or "male" e.g. might be considered as attributes that relate to the entity agent. In order to use ontologies in a way ontologies that not only represent entities, but also features of entities. do this follow this procedure:

1. Use the drag and drop labeled Concepts not forming edges at the bottom of P2, 10. Output Options index card. Click Refresh to automatically generate a list of all entities in the applied ontology.
2. Use the > and < buttons to move entities from one window to the other.
3. Apply your decision by hitting the Apply Selection button.
4. Before running Sub-Matrix Text Analysis decide whether you want to perform "Network Text Analysis" (NTA) or "Social Network Text Analysis" (SNTA). To select a type use the toggle button at the bottom of P2, 10. Output Options index card. Both types are a form of Sub-Matrix Text Analysis, but differ in what they measure:
  1. NTA: If one measures textual network, then entities of the applied ontology that should not be considered as nodes, but as inherent information of nodes, fill structural position and semantic function in a text. Thus they contribute to a texts' density, and therefore should be taken into consideration as statements for maps and stats. The resulting DyNetML file will contain nodes and inherent information on

nodes (if any contained in the window that an entity and an attribute co-occur).

2. SNTA: If one measures a social network that is represented in or extracted from a text, then links from entities of the applied ontology that should not be considered as nodes, but as inherent information of nodes, to actual other nodes are not to be represented in the maps and stats, since this would over fit the social network. The resulting DyNetML file will NOT inherent information on nodes.

**Note: Whether to use NTA or SNTA is a "text-philosophical question" and the answer depends upon what the user wants to measure: a textual network or a social network.**

Run Sub-Matrix Text Analysis.

Two options for creating DyNetML files are offered. Either one or both options can be selected per analysis:

1. Create one DyNetML file per map and text. To do this, check per Text in the Additional Output Formats field on P2, 10. Output Options index card. The DyNetML files will be stored as NameOfText.xml and in a folder called xml under the root directory of AutoMap.
2. Create one DyNetML file that unifies all maps. To do this, check per TextSet in the Additional Output Formats field on P2, 10. Output Options index card. The DyNetML file will be stored as consolidated\_map.xml in a folder called xml under the root directory of AutoMap.

### ***3.3 Examples for additional Output Formats***

Extracts from the Our Text I.txt and Our Text II.txt files were used as input:

Our Text 1.txt: Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

Our Text II.txt: Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray was pleased too. Mrs. Brown watered the roses and fertilized the garden on June 13, 1880. Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed how the plan had worked. John Darren and Mrs. Brown would take down the scarecrow the following week. She thought it was too scary for the dog. Mr. Craye painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.

Then AutoMap's Extensive Delete List, Direct Adjacency was applied to both texts. The Delete List was extended by further non-content bearing words that appeared in the sample texts (a an and as at awhile but for from happening he her her hers him his i in into it its me mine my nor of or our she so that the their theirs them they to us was we were what who whoever whom whomever will would you your yours yourself).  
Resulting texts:

Our Text I.txt: Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.  
Our Text II.txt: Mr. Cray's brown dog stopped eating lotus blossom 12 pm next day. Mrs. Brown now happy with dog. said "You good dog!" dog no longer ate flowers weeds. Mr. Cray pleased too.  
Mrs. Brown watered roses fertilized garden on June 13, 1880.  
Prof. Darren, Mrs. Brown & Mr. Cray met over dinner discussed how plan had worked. John Darren Mrs. Brown take down scarecrow following week. thought too scary dog. Mr. Craye painted fence. Then Mrs. Brown watered lotus, carnations, daffodils, roses.

Then map analysis was run on both texts using the AutoMap's default settings. All term distribution matrices provided on the Output Options panel were checked. Again, the default settings were used. The screen shot below shows the resulting analysis that should be displayed in P3.

Semantic network of current Text		
Frequency	Concept 1	Concept 2
5	mrs	brown
2	mr	cray
2	darren	mrs
2	brown	planted
1	you	impossible
1	yelled	saying
1	with	dog
1	weeds	asked
1	weeding	was
1	weeded	garden
1	was	unhappy
1	vole	on
1	up	scarecrow
1	unhappy	with
1	thought	scare
1	then	mrs
1	stop	dog
1	solved	then

## 4. Network Analytic Measures

AutoMap supports the computation of network analytic measures per map or network and per concept or node.

**Note: Only Network Analytic Measures for directed networks were implemented into AutoMap. The reason for this is that AutoMap outputs are always directed in order to adequately represent the linear structure of texts.**

To create measures follow these steps:

1. On P2, tab no. 10 Output Options, select the Network analytic measures field and check Measures.
2. After analysis is run, measures will be displayed on P4, 2. Network analytic measures index card.  
A file that collects all measures per text set will be saved as overall\_network\_analytix\_measures.csv in the results folder under the root directory of AutoMap.

**Note: Texts have a linear structure. Therefore, we only implemented Network Analytic Measures for directed networks (digraphs) into AutoMap.**

The following Table explains the measures that can be computed:

Name of measure	Calculation of measure	Name of and reference for corresponding SNA measure	Corresponding name of measure in Cube
<b>Concept (node) level measures, direct connectivity</b>			
<b>Concept outdegree</b>	Total number of statements with concept in anterior position	Outdegree, expansiveness, actor degree centrality (Wasserman & Faust 1994); Prestige, Influence (Mrvar)	Local imageability
<b>Concept indegree</b>	Total number of statements with concept in posterior position	Indegree, receptivity, popularity, actor degree prestige (Wasserman & Faust 1994); Prestige, Support (Mrvar)	Local evokability
<b>Concept outdegree centrality</b>	Total number of statements with concept in anterior position/ Number of unique concepts in text	Outdegree Centrality (Wasserman & Faust 1994); Relative Influence (Mrvar)	None
<b>Concept</b>	Total number of	Indegree Centrality	None

<b>indegree centrality</b>	statements with concept in posterior position/ Number of unique concepts per text	(Wasserman & Faust 1994); Relative Support (Mrvar)	
<b>Total degree</b>	Concept indegree + concept outdegree	-	Local density
<b>Map (graph) level measures, direct connectivity</b>			
<b>Text outdegree centrality</b>	Sum (largest observed outdegree – outdegree of concepts)/( Number of unique concepts in text) 2	Group outdegree centralization (Wasserman & Faust, 1994)	None
<b>Mean concept outdegree centrality</b>	Sum (outdegree)/ Number of unique concepts in text	Mean outdegree (= Mean indegree) (Wasserman & Faust 1994)	None
<b>Variance of concept outdegree centrality</b>	Sum(sum outdegree – mean outdegree) 2 / Number of unique concepts in text	Variance of outdegree (Wasserman & Faust 1994, p.127-128)	None
<b>Text indegree centrality</b>	Sum (largest observed indegree – indegree of concepts)/( Number of unique concepts in text) 2	Group indegree centralization (Wasserman & Faust, 1994)	None
<b>Mean concept indegree centrality</b>	Sum (indegree)/ Number of unique concepts in text	Mean indegree (=Mean outdegree)	None
<b>Variance of concept indegree centrality</b>	Sum (sum indegree – mean indegree) 2 / Number of unique concepts in text	Variance of indegree (Wasserman & Faust 1994, p.127-128)	None
<b>Density</b>	Number of statements/ Possible number of statements	Density (Wasserman & Faust 1994, p.129, Scott 1991, p.74) Wasserman and Faust use $g(g-1)$ as denominator; we use $(g*g)$ , because unique concept can form statement with same unique concept (e.g. agent-agent)	None
<b>Concept (node) level measures, indirect connectivity</b>			
<b>Concept closeness centrality</b>	Minimum possible total distance from node i to all other nodes/ Sum of all geodesics between node i and all other nodes	Closeness (Wasserman & Faust, Mrvar) According to Wasserman and Faust (1994, p.200) group level closeness centrality is not computed	None

<b>Concept betweenness centrality</b>	sum ((Number of all geodesics between all nodes that go through node i )/( Number of geodesics between node i and all other nodes))/(( Number of unique concepts in text - 1)(Number of unique concepts in text -2))	Betweenness (Gould 1987, Mrvar)	None
<b>Concept proximity prestige</b>	Number of concepts directly or indirectly adjacent to node i	Proximity Prestige (Wasserman & Faust 1994, Mrvar)	None
<b>Map (graph) level measures, indirect connectivity</b>			
<b>Text Proximity Prestige</b>	Sum (Proximity Prestige (all unique concepts in text)/ Number of unique concepts in text	Group level proximity prestige (Wasserman & Faust 1994)	None

## Entropy

Entropy is the formalization of redundancy and diversity. Thus we say that Information Entropy (H) of a text document (X) where probability p of a word x = ratio of total frequency of x to length (total number of words) of a text document.

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$$

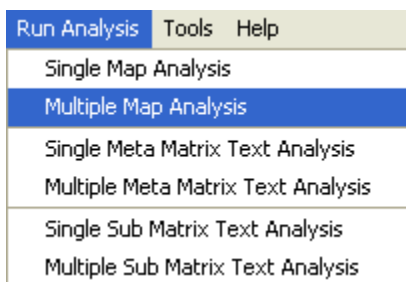
To run Entropy you will need to load multiple text documents into AutoMap for analysis.

### File > Open Multiple Files > Select Folder.

Navigate to the folder containing the texts you wish to analyze, and single click on it.

In the menu bar, select option Run Analysis

Select the Multiple Map Analysis option as shown in the diagram below:





Once you have selected the Multiple Map Analysis, go to Panel 2, tab no. 4. Analysis Settings, then select Tab no. 2. Output Options.

You will find the information in an excel file, which will be located in the last column of the stats\_overall.csv output file.

## ***Reference***

Theory:

*W. Weaver & C. E. Shannon, The Mathematical Theory of Communication, Urbana, Illinois: University of Illinois Press, 1949*

Empiric application:

*Diesner, J., Carley, K.M., & Katzmair, H., The Morphology of a Breakdown: How the semantics and mechanics of communication networks from an organization in crises relate. XXVII Sunbelt Social Network Conference, Corfu, Greece, May 2007.*

## DYNetML

### ***Changes in DyNetML outputs for XML files per Text***

The DyNetML outputs per Map (Semantic Network Analysis) and per Text (Sub-Matrix Analysis) have been changed to reflect:

- The new DyNetML specification (<http://casos.isri.cmu.edu/dynetml/index.html>)
- Isolated entities
- Frequency of entities
- A set of attributes on the data and statistical results

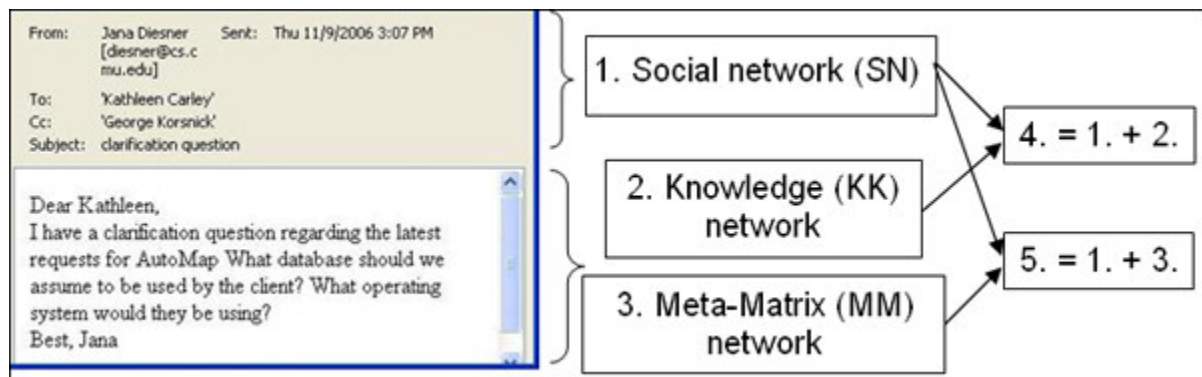
This change does not impact the consolidated DyNetML files as produced in AutoMap.

# CASOS Email Parser (CEMAP)

## 1. Instructions for the using the CASOS Email Parser (CEMAP) through AutoMap

### Introduction

The CASOS Email Parser (CEMAP) that is launched through AutoMap (<http://www.casos.cs.cmu.edu/projects/automap/software.html>) enables the extraction of different types of network information from emails (e.g. who exchanges information, who provides what information, etc.). The following image shows what types of information can be extracted with CEMAP.



Network type1 (Social network) represents social network data that can be extracted from email headers (e.g. agent-agent networks). In SN, entities represent people, and edges represent exchanged emails (frequency count). This network type does not require any text coding in AutoMap.

Networks types 2 and 3 are extracted by performing text coding in AutoMap. More specifically, Knowledge networks (KK) represent semantic network or mental models that are contained in the bodies of individual emails. In KK, entities represent knowledge items, and edges represent the co-occurrence of terms in text. For MM, texts are coded in AutoMap according to a taxonomy or ontology (e.g. meta-matrix, while are ontologies can be specified by the user). In MM, entities represent instances of categories (e.g. agent, knowledge, resources) of the ontology, and edges represent co-occurrences of terms in texts.

Networks types 4 and 5 result from the combination of SN with KK and MM, respectively. In type 4, entities represent people and knowledge, and edges represent emails and mental models. In type 5, nodes represent the categories of the taxonomy as specified by the user, and the edges represent the co-occurrence of the terms that represent instances of the taxonomy in the corpus. For the creation of type 4 and 5, the extraction of type 2 and 3, respectively, is mandatory.

CEMAP stores all network data as DyNetML files (a derivate of XML). This data can be analyzed with any package that reads DyNetML. We recommend using the ORA software for this purpose (<http://www.casos.cs.cmu.edu/projects/ora/index.html>).

## **How to do it:**

### **1. Extraction of network type 1 data:**

- 1.1. This network type is needed to perform social network analysis (SNA) based on header information only.
- 1.2. In CEMAP, Analysis Type, select the type of your mail box.
- 1.3. In CEMAP, Input Data Set, specify your email client, inbox type, username, and password as semicolon separated values, e.g. `pop.gmail.com;INBOX;JohnDoe;JohnDoesPassword`
- 1.4. In CEMAP, Enter Destination Directory for Texts, put in a directory of your choice, e.g. `C:\textInput`. The email bodies will be stored in AutoMap format (.txt) under this directory. Make sure that this directory does exists on your machine.
- 1.5. In CEMAP, Enter Destination File for Network, put in a directory AND filename of your choice. The filename extension needs to be .xml. e.g. `C:\output\AA.xml`.
- 1.6. You are now ready to extract network type 1. In order to do so, press the Start Extraction button. As a result, CEMAP will extract network data from the email headers (e.g. agent by agent) and store it as one XML file under the directory that is specified in the Destination File for Network.
- 1.7. You can pick up the output file and load it into the SNA package of your choice. We recommend using the ORA package (<http://www.casos.cs.cmu.edu/projects/ora/download.php>).

To extract networks of type 2, perform Semantic Network Analysis in AutoMap.

To extract networks of type 3, perform Meta-Matrix Text Analysis in AutoMap.

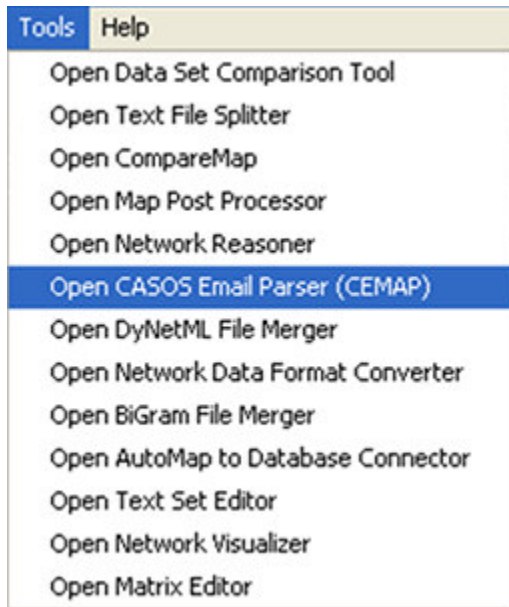
To extract networks of type 4, merge networks of type one and 2.

To extract networks of type 5, merge networks of type one and 3.

CEMAP is a feature in AutoMap that will import email files into AutoMap to analyze the text in the emails, such as who exchanges information, who provides what information, etc.

## **To use the feature:**

- In AutoMap menu bar, select Tools > Open CASOS Email Parser (CEMAP).



Once that is selected the CASOS — Email Analyzer window will appear. The Email Analyzer window is where you will type in your information of what type of network you would like to create while importing your emails.

#### **The different types of imports are**

- Full email
- Header information only
- Forwarded/New emails
- Subject lines only

Select the type of Format your email is in, from the drop down menu.

- MBOX — File that is stored in standard file folders
- POP3 and IMAP — Emails that are on a server
  - Will require Username, User password, Server Name, and Folder Name
- TXT — Exported email files from outlook
  - Will require a Data (file) set identifier
- Enron and CasosDB are for internal use only
- Dynetml — Currently under development

Once you have selected what type of format you would like to use, click on the Start Extraction button. This will begin the process of importing your emails.

## Support

### ***How to cite AutoMap***

*Carley, K.M., & Diesner, J. (2005). AutoMap: Software for Network Text Analysis.*

### ***Further Information***

For further information on AutoMap please visit:

<http://www.casos.cs.cmu.edu/projects/automap>

On this web page you will find:

- Publications about AutoMap, Network Text Analysis, and Map Analysis
- Installers for AutoMap
- Contact information
- Information about sponsors of AutoMap
- Information about people working on AutoMap

We also provide a online discussion forum for AutoMap:

- to discuss questions related to the AutoMap software;
- and get help from other AutoMap users and the developers of the software with using the tool.

### ***Questions, Bugs, and Comments***

Please contact us:

Dr. Kathleen M. Carley (kathleen.carley@cmu.edu)

Jana Diesner (diesner@cs.cmu.edu)

Technical Writer, Matt De Reno (mjdereno@andrew.cmu.edu)

Carnegie Mellon University

School of Computer Science

Institute for Software Research International (ISRI)

Center for Computational Analysis of Social and Organizational Systems (CASOS)

5000 Forbes Avenue

1325 Wean Hall

Pittsburgh, PA, 15213

We provide an online discussion forum for AutoMap:

- To discuss questions related to the AutoMap software.
- And get help from other AutoMap users and the developers of the software with using the tool.

# Lesson 1

## *From content analysis to semantic networks*

### *I. Open AutoMap*

#### *On Empire:*

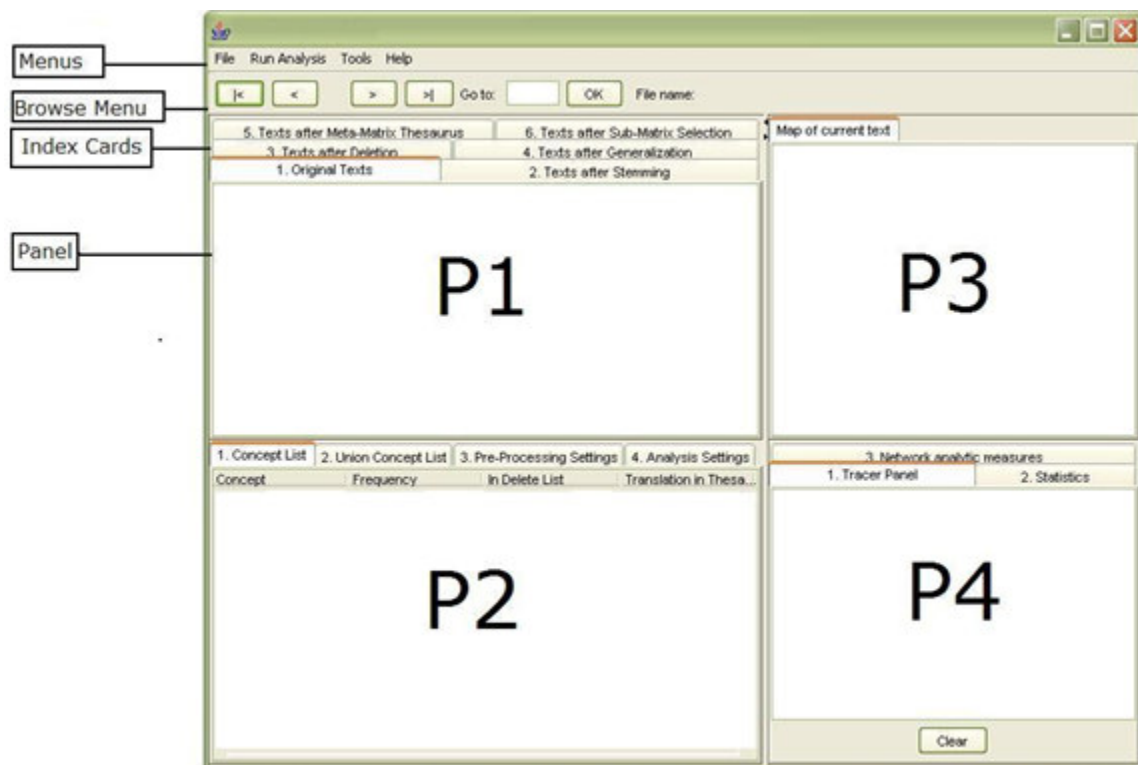
- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run Automap icon. The Automap Graphical User Interface (GUI) pops up.

**From anywhere else:** (requires a web connection)

<http://www.casos.cs.cmu.edu/projects/automap/software.html>

### *II. Overview of the AutoMap Graphical User Interface (GUI)*

Below is a screen capture of the AutoMap GUI:



The AutoMap GUI is divided into four primary quadrants, as shown above, or panels: P1 (top left), P2 (bottom left), P3 (top right), and P4 (bottom right).



A drop-down menu bar at the top of the window will provide access to analysis tools and utilities. The Browse Menu allows you to quickly navigate between texts you have loaded into AutoMap.

Panel P2 can be edited, but the remaining panels (P1, P3, P4) cannot. The information that is displayed on each panel is always related to the information displayed in the other three panels. The Browse Menu, however, relates to all four panels at the same time.

### ***III. Load multiple text documents into AutoMap***

In order to analyze multiple pieces of text simultaneously, you will need to save the documents together in one folder.

#### **Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\ TextInputFiles**

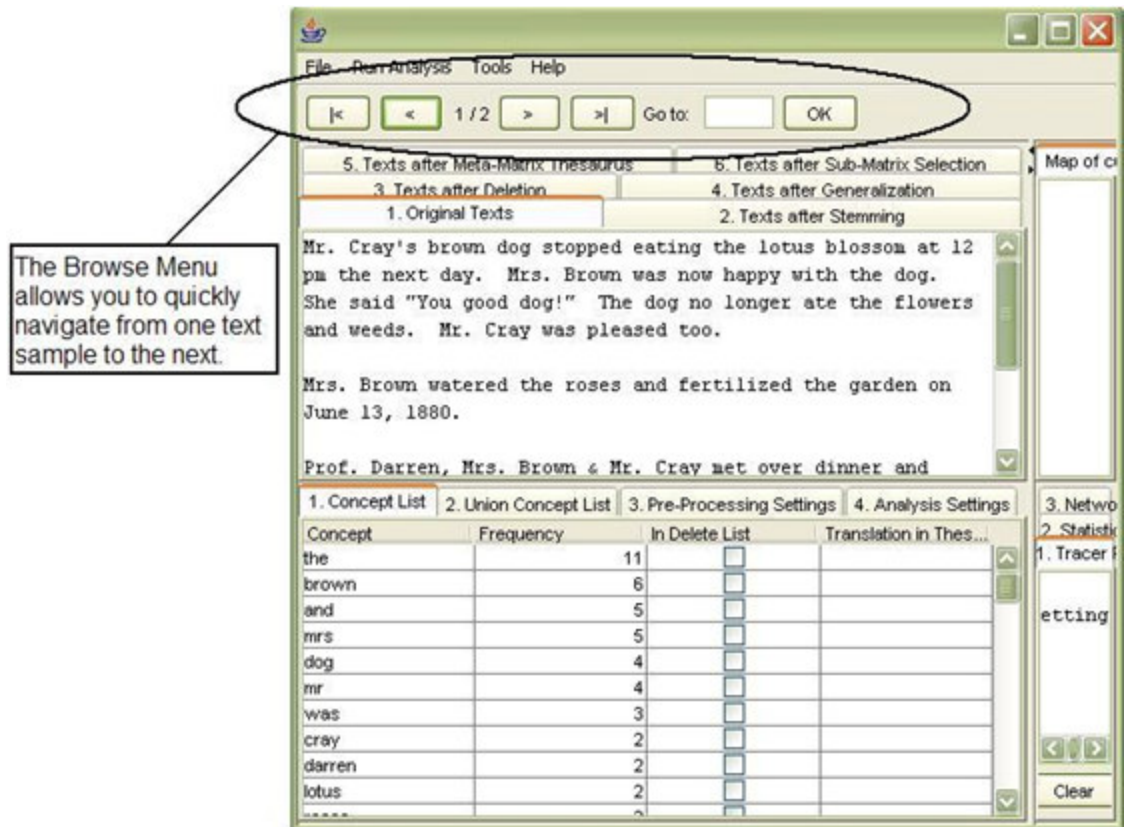
- In the AutoMap menu bar, select File > Open Multiple Files. A folder selection window will pop up.
- Navigate to the folder containing the documents you wish to analyze, and single-click on it.

#### **Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\ TextInputFiles**

- Click the Open button in the bottom right corner.

The first text will be displayed in panel P1 under the tab titled 1.Original Texts. You can use the Browse Menu to navigate through the texts.

Below is a screen capture of the first text display:



#### IV. Carry out a simple content analysis

A simple content analysis determines the frequencies of all words in a text document.

##### Examine the Concept List:

The Concept List considers concepts in each text file individually. It is created automatically when a document or set of documents is loaded into AutoMap. This will be displayed in the P2 panel under the tab titled 1. Concept List.

The Concept List will tell you several important facts about each document in your set:

- The number of concepts found in the text displayed in P1 under the tab titled 1.Original Text
- Related frequencies of those concepts
- Cumulated unique concepts and total concepts contained in the data set

**Note: The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts.**

By default, the Concept List sorts by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To resort the list by frequency, click on the second-column header Frequency.

Below is a screen capture showing part of the Concept List, sorted by frequency, for the first document in our loaded folder of texts (Text1-in-user's guide.txt):

Click this column heading to sort concepts alphabetically

Click this column heading to sort concepts by frequency

Concept	Frequency	In Delete List	Translation in Thes...
the	10	<input type="checkbox"/>	
brown	6	<input type="checkbox"/>	
dog	6	<input type="checkbox"/>	
mrs	5	<input type="checkbox"/>	
a	4	<input type="checkbox"/>	
and	4	<input type="checkbox"/>	
mr	4	<input type="checkbox"/>	

### ***Create and refresh the Union Concept List:***

The Union Concept List considers concepts in all texts in a set (unlike the Concept List, which considers each text in the set individually). It is located in Panel P2

under the tab titled 2.Union Concept List.

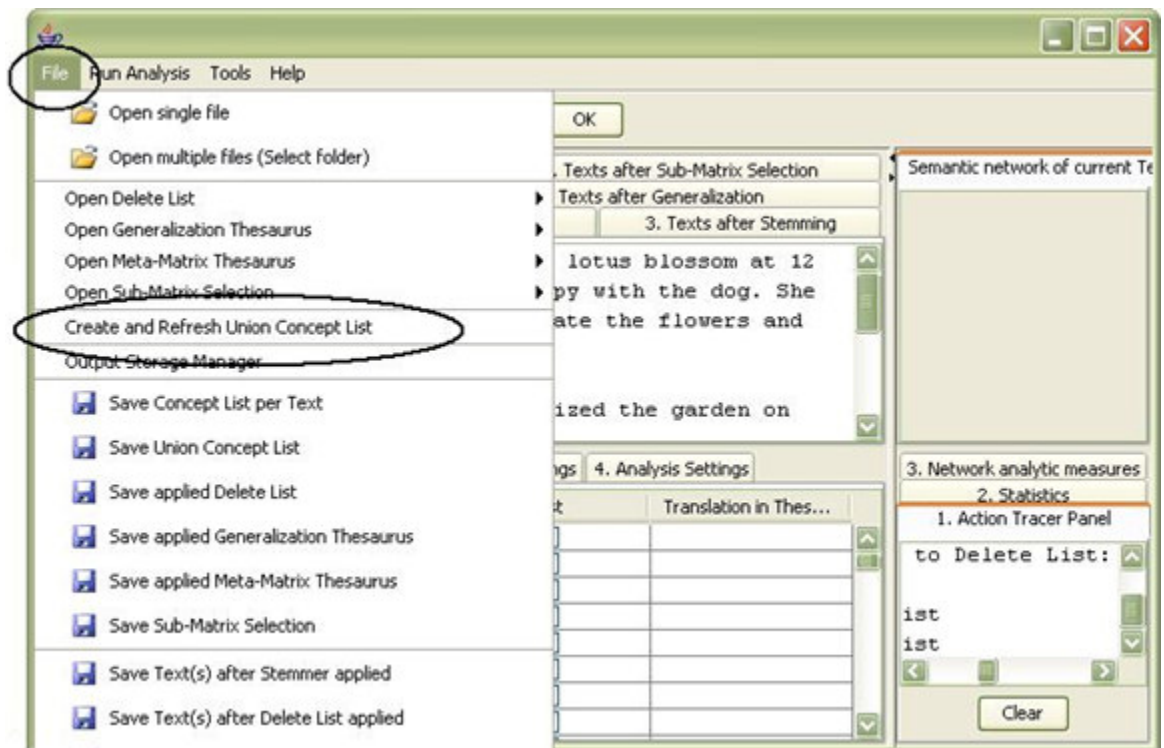
The Union Concept List tells you several important facts about your text set:

- Concepts contained in all loaded text sets
- Related, cumulative frequencies of concepts in all text sets
- Cumulative unique and total concepts

Before viewing your Union Concept List, you must refresh it. To refresh your Union Concept List, go to the Menu Bar, choose File > Refresh Union Concept List.

The Union Concept List can be refreshed after each step of pre-processing. This will allow you to visualize the impact of pre-processing operations on the union of concepts.

Below is a screen capture showing where to find the Refresh Union Concept List command:



After refreshing your Union Concept List, you can view it under the 2.Union Concept List tab in Panel P2. Below is a screen capture showing where this tab is located in the GUI:

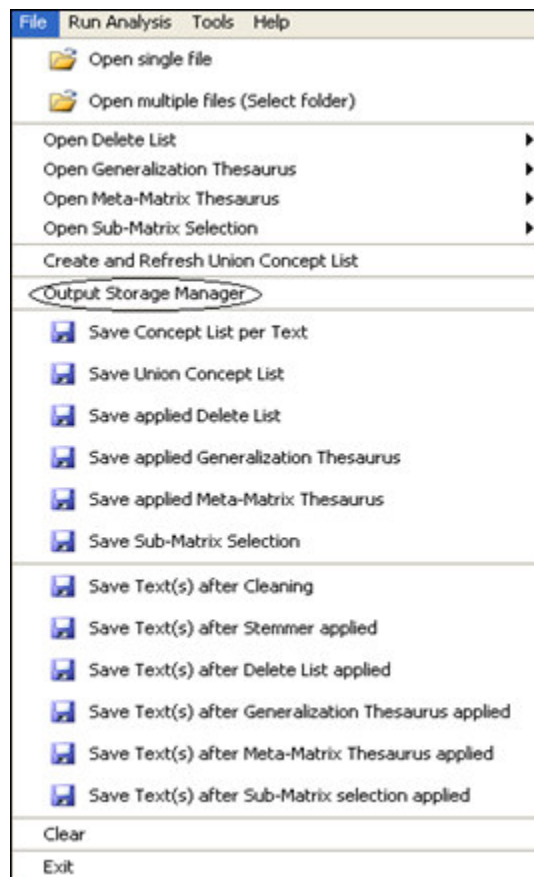
By default, the list is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To re-sort the list by Frequency, click on the second-header column Frequency.

**Note: The number of unique concepts considers each concept only once. The number of total concepts also considers repetitions of concepts.**

## ***V. Save your Union Concept List as a csv file***

To specify a location for your Union Concept List file, choose File > Output Storage Manager.

Below is a screen capture showing where to find this in the GUI:

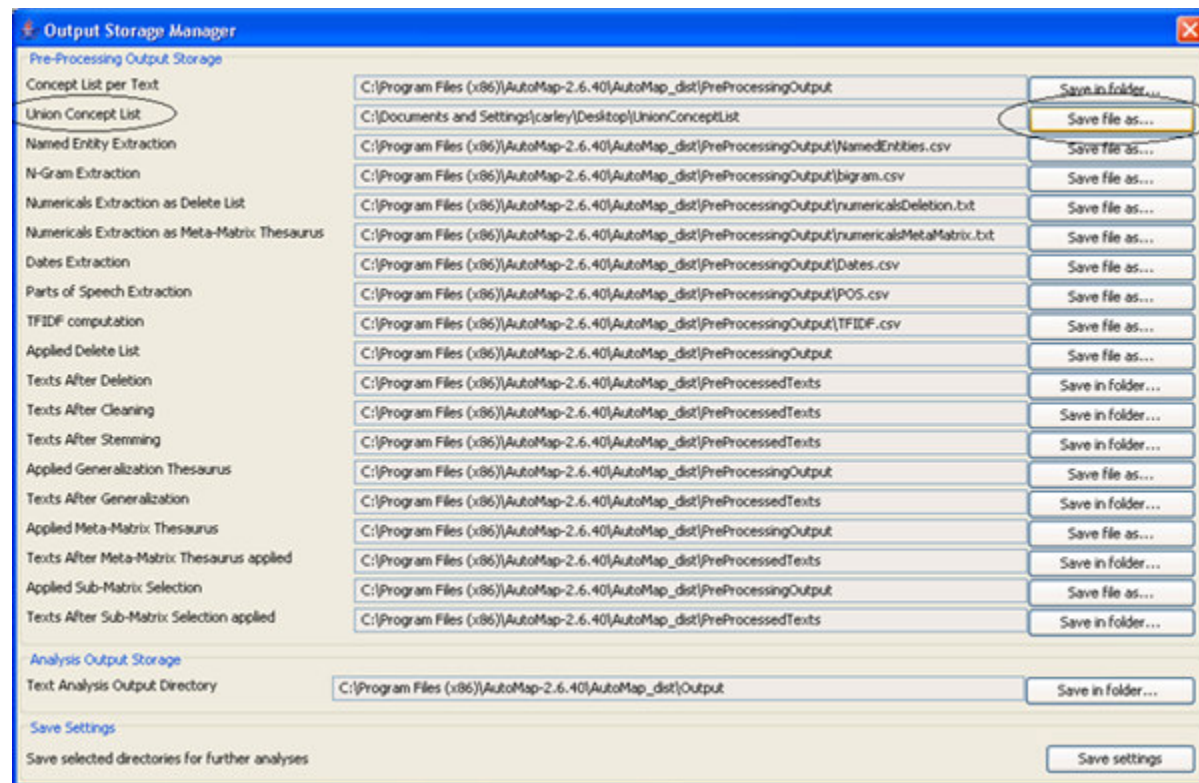


A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

The Union Concept List appears as the second item in the Output Storage Manager. Click the Save file as... option to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want.

## Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\UnionConceptList

Below is a screen capture showing where to find the correct Save file as... button in the Output Storage Manager:



Click the Save Settings button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the red X at the top right corner.

In the Automap Menu Bar, choose File > Save Union Concept List.

## VI. View your csv file in Microsoft Excel

Navigate to the location of your new csv file and double-click on its icon.

## Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\UnionConceptList

A box titled “Open With” pops up. Click on Microsoft Excel, then click OK.



**Note: You must have MS Excel installed on your computer in order to view your csv file this way.**

### ***VII. Save and view your results as a binary csv file***

Follow Steps 3 and 4.

### ***VIII. Carry out a simple Semantic Network Analysis***

A Simple Network Analysis involves no thesauri and uses all words. A Semantic Network Analysis can be run on original input texts or on texts that have been pre-processed with a Delete List and/or a Generalization Thesaurus.

#### **Before running the analysis:**

Make sure you have completed the following steps:

- Pre-process the texts if you wish
- Specify the analysis settings (otherwise the default settings will be applied)
- Specify additional output options if you wish to (if not, the default outputs will be generated)

#### **About the analysis settings:**

Use the Analysis Settings index card to specify the analysis settings. Your settings will be automatically applied to the analysis (you do not need to confirm them).

Below is a screen capture showing where to find the Analysis Settings index card in P2:

The screenshot shows the 'Analysis Settings' tab in the AutoMap software. It is divided into two sub-tabs: '1. Analysis Settings' and '2. Output Options'. The 'Analysis Settings' sub-tab is active. It contains two main sections: 'Coding Ties' and 'Windowing'. In the 'Coding Ties' section, under 'Directionality', the 'Uni-directional' radio button is selected. Under 'Strength', the 'Frequency' checkbox is checked. In the 'Windowing' section, under 'Punctuation', the 'Ignore punctuation completely' radio button is selected. Under 'Window-Size', the value '2' is selected from a dropdown menu. A button labeled 'Find size of largest text in text set' is located at the bottom right of the 'Windowing' section.

The following chart lists the possible analysis settings:

<p><b>Coding Ties</b> Specify the way statements are counted.</p>	<p><b>Directionality</b> Select one of the following two possibilities by checking the button.</p>	<p><b>Uni-Directional</b> (When coding a tie, only 1st-&gt;2nd concept should be noted) <b>Bi-Directional</b> (When coding a tie, both 1st &lt;-&gt; 2nd concept shall be noted)</p>
<p><b>Windowing</b> Windowing is a method that codes the text as a map by placing relationships between pairs of Concepts that occur within a window.</p>	<p><b>Strength</b> Strength will be printed out by default. To not print out Strength uncheck the item.</p>	<p><b>Frequency</b> (The cumulative frequency of every existing statement.) <b>Item not checked:</b> Existence of frequency will be printed out (binary result).</p>
	<p><b>Punctuation</b> Select one option by checking the radio button.</p>	<p><b>Ignore punctuation completely</b> (Statements will be placed between all concepts.) <b>Reset window at end of paragraphs only</b> (Statements will be placed only within every single paragraph.)</p>
		<p><b>Reset window at end of paragraphs and sentences</b> (Statements</p>



will be placed only within every single sentence.)

Window size between **2 and 100**.

The Window Size defines how distant concepts can be and still have a relationship. Only concepts in same window can form statements.

### **Window Size**

Select one window size by using the number chooser.

If you do not want to change any of the suggested options, the analysis will be done with a set of standard (default) settings.

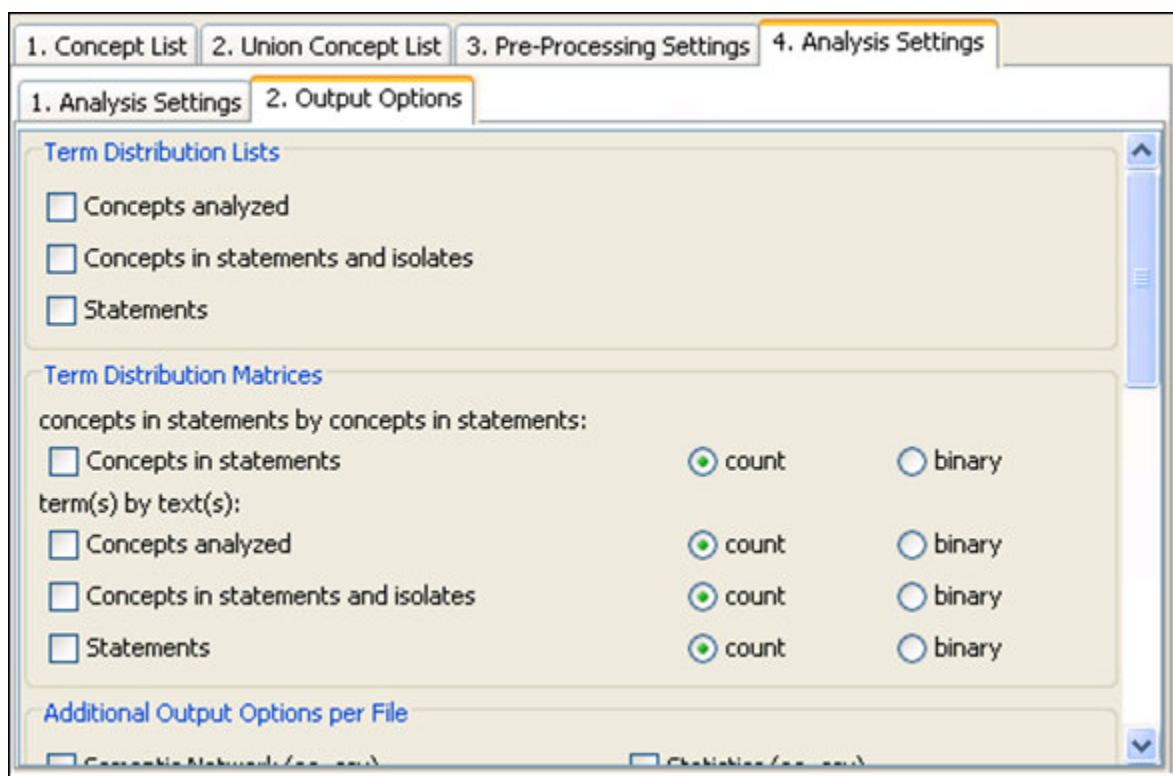
Below is a screen capture showing the standard settings:

The screenshot shows the '4. Analysis Settings' tab in the AutoMap interface. It contains two sub-tabs: '1. Analysis Settings' and '2. Output Options'. Under '1. Analysis Settings', there are two sections: 'Coding Ties' and 'Windowing'. In the 'Coding Ties' section, 'Directionality' has 'Uni-directional' selected (indicated by a green dot) and 'Bi-directional' is unselected. 'Strength' has 'Frequency' unselected. In the 'Windowing' section, 'Punctuation' has 'Ignore punctuation completely' selected (indicated by a green dot), 'Reset window at the end of paragraphs only' is unselected, and 'Reset window at the end of paragraphs and sentences' is unselected. The 'Window-Size' is set to '2' in a dropdown menu. A button labeled 'Find size of largest text in text set' is located at the bottom right of the 'Windowing' section.

### **About the output options:**

The map and the statistic output generated by AutoMap are displayed in Panels P3 and P4, respectively. In addition, AutoMap offers further output options that can be selected in the 10.Output Options index card in Panel P2. Any other additional outputs are generated after analyses are run.

Below is a screen capture showing the 10.Output Options index card in P2:



For all types of multiple analysis, term distribution lists and matrices can be selected as output options in the upper two fields of the 10.Output Options index card in P2.

### Points to consider:

- No list or matrix is generated by default.
- Lists or matrices are only generated if the user checks the item wanted and runs an analysis (of any type).
- If pre-processing was performed, the list relates to the stage of pre-processing that was used for the analysis. If several pre-processing techniques are applied, analysis will run on the highest stage of pre-processing.
- The requested lists and matrices are automatically saved in a folder called Term Distribution Lists and Matrices under the root directory of AutoMap. This folder is overwritten with new analysis you run. If you want to save the results somewhere other than the default Term Distribution Lists and Matrices folder, simply rename that folder.
- Two output lists are generated for each Term Distribution List checked.
- One output matrix is generated for each Term Distribution Matrix checked.

The following chart lists the types and content of Term Distribution Lists and Matrices:

Output Type	Name of output	Content of output	
		List of concept analyzed.csv	Concept, Text, Frequency
Term Distribution List	Concepts analyzed		Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts
		Statistics of concepts analyzed.csv	
		List of concept in statements.csv	Concept, Text, Frequency
		Statistics of concept in statements.csv	Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts
	Concepts in statements and isolates	List of isolates in statements.csv	Concept, Text, Frequency
		Statistics of isolates.csv	Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts
		List of statements.csv	Statement, Text, Frequency
		Statistics of statements.csv	Statement, Cumulated sum across text set, Number of text statement occurs in, Percentage of texts statement occurs in, Texts
Term Distribution Matrices	Concepts in statements by concepts in statements:		
	Concepts in statements	Matrix of concepts in statements.csv	Matrix of Concept that were linked into statements (first row) by Concept that were linked into statements (first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence

Term(s) by text(s):		(1) or absence (0) of concept
Concepts analyzed	Matrix of concept analyzed.csv	Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
Concepts in statements and isolates	Matrix of concept in statements.csv	Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
Statements	Matrix of statements.csv	Matrix of Statements (union of statements listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of statement If binary was chosen, cells denote existence (1) or absence (0) of statement

### **Save your upcoming analysis in DyNetML format:**

Before running your Multiple Map Analysis, you must specify that your results will be in DyNetML format so that the results can eventually be read into ORA.

On the 10.Output Options index card in Panel P2, scroll down to the option titled “Additional Output Formats” and check the per Map box next to “DyNetML for Map Analysis.”

Open the Output Storage Manager. (See section V.)

You will see “Text Analysis Output Directory” listed under the “Analysis Output Storage” category. Click the Save in folder... button on the far right in that row. In the folder chooser window that pops up, navigate to the folder you want.

**Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextOutputFiles**

Click Open.

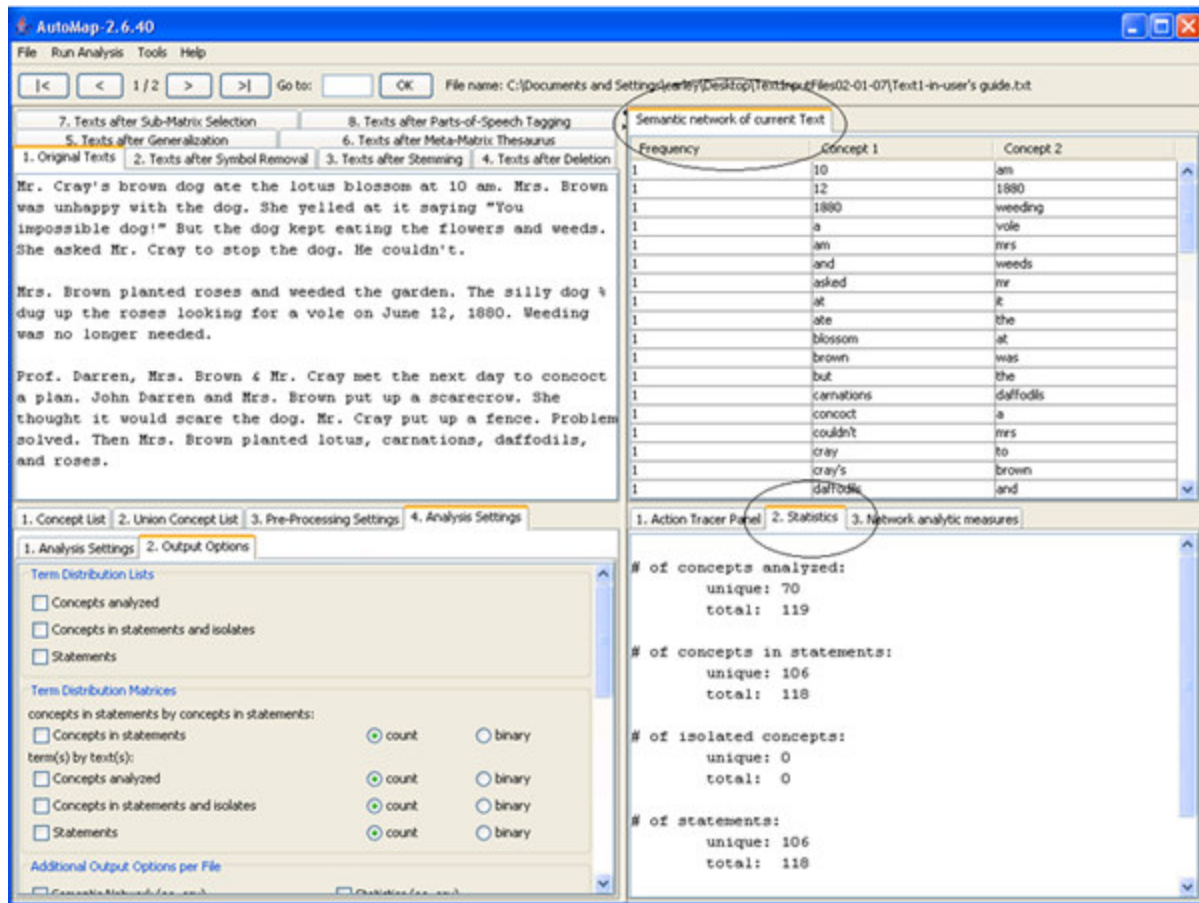
Save these settings and close the Output Storage Manager. (See section V.)

### **Run the Multiple Map Analysis:**

In the Menu Bar, choose Run Analysis > Multiple Map Analysis.

The results will be displayed on the Semantic Network of Current Text index card in Panel P3 and on the 2.Statistics index card in Panel P4.

Below is a screen capture showing the results of Multiple Map Analysis in Panels P3 and P4:



If you have analyzed multiple texts, you can browse through the results and see the related texts. Results of multiple analyses are automatically saved in a folder titled “output” under the directory where Automap is installed.

The output folders are overwritten with every new analysis you run. If you want to save the results of a current analysis folder, simply rename that folder.

## IX. Open ORA

### On Empire:

- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run ORA icon. The ORA interface pops up.

**From anywhere else:** (requires a web connection)

<http://www.casos.cs.cmu.edu/projects/ora/software.html>

## ***X. Load your files into ORA***

Click the Load button on the far right. A file chooser window will pop up.

Choose Select Files. Navigate to your NyNetML files (generated in section VIII).

**Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextOutputFiles**

Select the files you want by holding down the <ctrl> key while clicking on each one. Click Open, then Finish.

**Example:**

**15. Select Text1-in-user's guide.txt**

**16. Press the Control key**

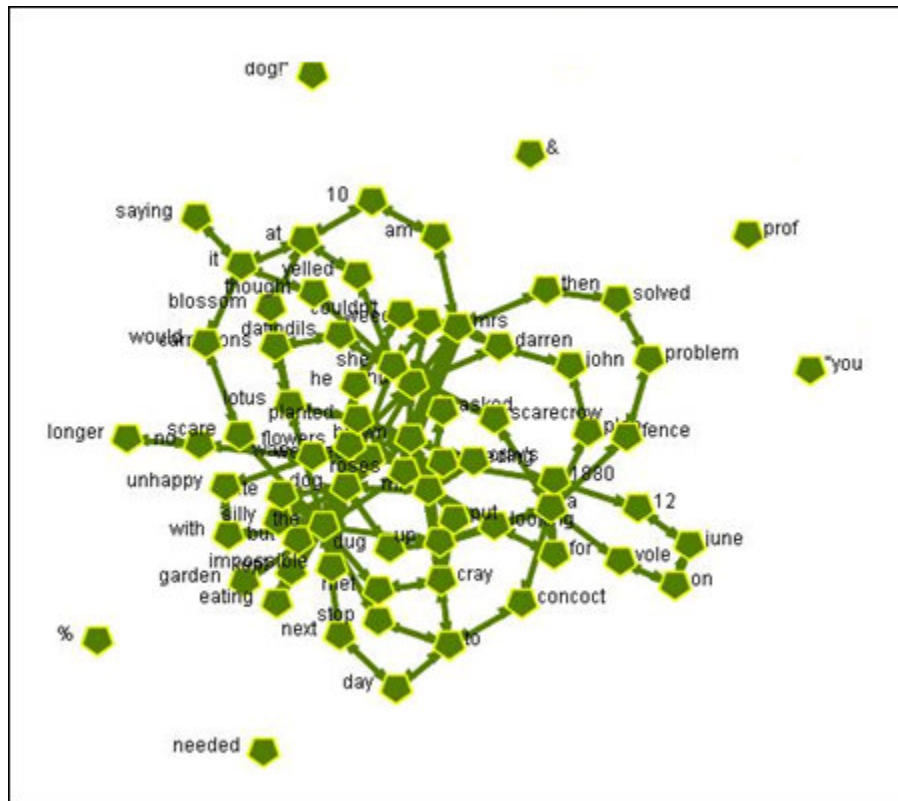
**17. Select Text2-in-user's guide.txt**

**Note: If some file is already loaded into ORA, choose the Append As Additional Meta-Matrix option.**

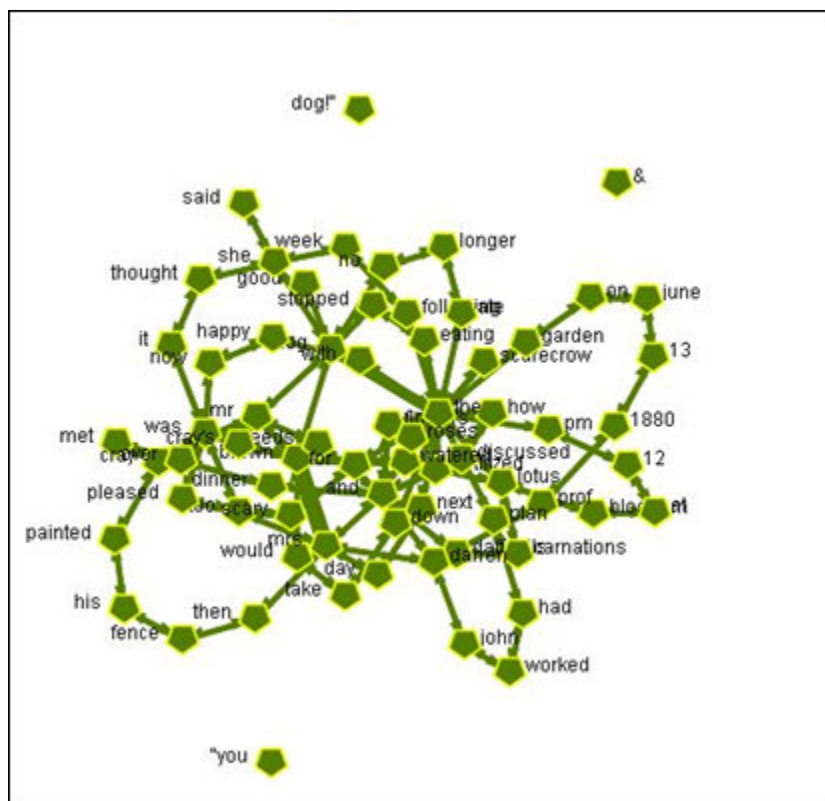
## ***XI. Visualize your meta-matrices***

The left-hand portion of the ORA screen is the ORA navigator. In this navigator, select the matrix you want to visualize by clicking on it. Click the Visualize this meta-matrix button near the center of the screen to generate the visualization. Below are the visualized meta-matrices for Text1-in-user's guide.txt and Text2-in-user's guide.txt:

Text1-in-user's guide.txt



Text2-in-user's guide.txt



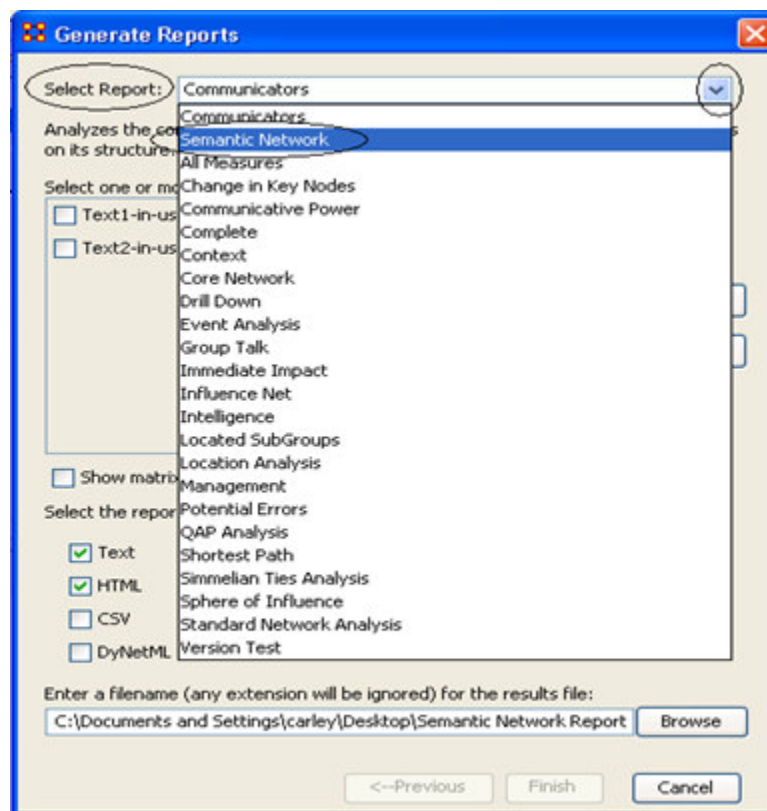


These meta-matrices also can be found in:  
C:\Documents and  
Settings\carley\Desktop\AutomapLesson1\MetaMatrices

## ***XII. Run a Semantic Network Report***

In ORA's Menu Bar, choose Analysis > Generate Reports. A window will pop up. You will see “Select Report” at the top of the window. Click the v-shaped icon to the right, and select Semantic Network from the resulting drop-down menu.

Below is a screen capture showing where to find the Semantic Network Report in the Generate Reports pop-up window:



Check the two boxes for Text1-in-user's guide.txt and Text2-in-user's guide.txt inside the field titled “Select one or more meta matrices”.

To specify a filename for the results, click the Browse button near the bottom right corner of the Generate Reports pop-up window. Navigate to the desired location and type in a filename or you can use the default name.

**Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\SemanticNetworkReport.html**

Click the Next button, then the Finish button at the bottom of the Generate Reports pop-up window.

ORA will run. An HTML file will pop up displaying the results.

Return to the ORA window. When a small box titled “Complete” has popped up, click the OK button.

Below is the Semantic Network Report generated for Text1-in-user's guide.txt and Text2-in-user's guide.txt:

## ***SEMANTIC-NETWORK REPORT***

Input data: Text1-in-user's guide.txt, Text2-in-user's guide.txt

Start time: Tue Mar 06 10:59:33 2007

This is a comparison of two semantic networks. Each node in the network is considered a Concept, and each edge a Statement connecting two concepts. Statement weights are interpreted as the number of times the statement occurred in the underlying input text.

<b>Network</b>	<b>Concepts</b>	<b>Statements</b>	<b>Density</b>
Text1-in-user's guide.txt	74	200	0.0370233
Text2-in-user's guide.txt	69	184	0.0392157
Union	98	310	0.032611
Intersection	45	8	0.0040404

## ***Symmetric difference***

The symmetric distance of network A to network B is a new network that contains the entities in A that are not in B.

Statements

<b>Network</b>	<b>Concepts</b>
Text1-in-user's guide.txt to Text2-in-user's guide.txt	192
Text2-in-user's guide.txt to Text1-in-user's guide.txt	176

## ***Saved output networks***

The summary statistics across all maps saved to: C:\Documents and Settings\jmcgille\_summary\_statistics.csv

Produced by ORA developed at CASOS - Carnegie Mellon University

## **Lesson 2**

### ***Processing data: using the delete list***

#### ***I. Open AutoMap***

##### **On Empire:**

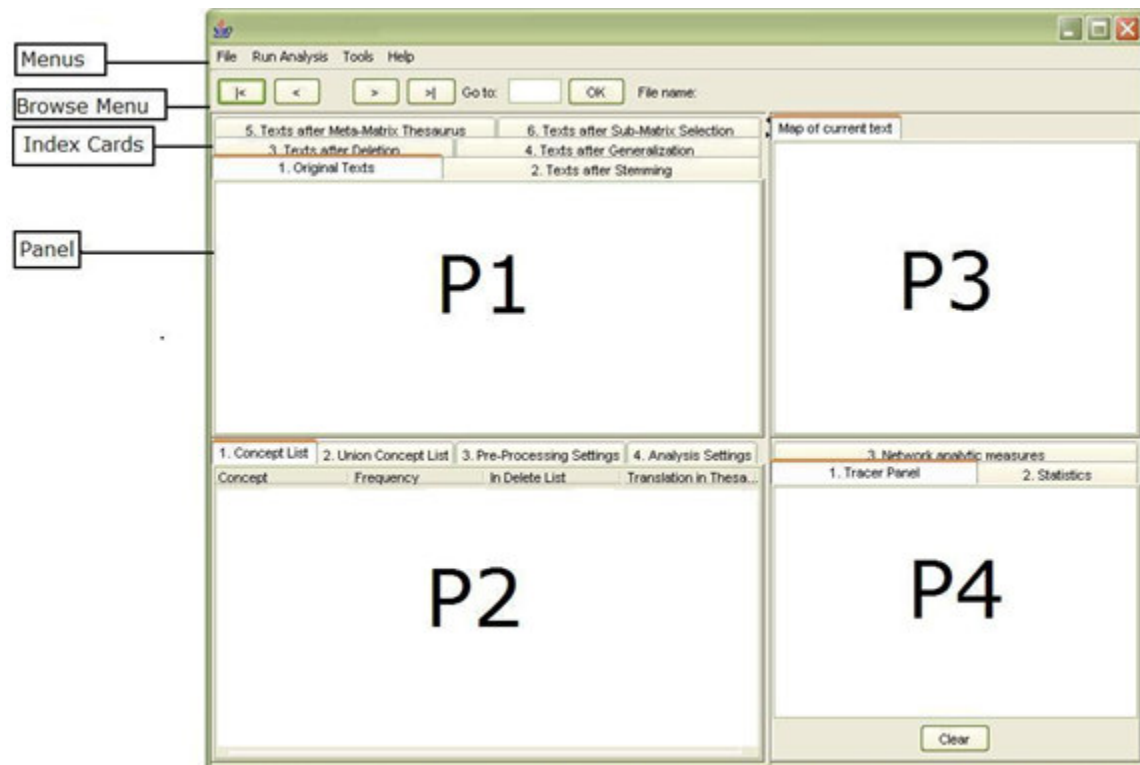
- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run Automap icon. The Automap Graphical User Interface (GUI) pops up.

**From anywhere else:** (requires a web connection)

<http://www.casos.cs.cmu.edu/projects/automap/software.html>

#### ***II. Overview of the AutoMap Graphical User Interface (GUI)***

Below is a screen capture of the AutoMap GUI:



The AutoMap GUI is divided into four primary quadrants, or panels: P1 (top left), P2 (bottom left), P3 (top right), and P4 (bottom right).

A drop-down menu bar at the top of the window provides access to analysis tools and utilities. The Browse Menu allows you to quickly navigate between texts you have loaded into Automap.

The P2 panel can be edited, but the other remaining panels (P1, P3, P4) cannot. Information displayed on each panel is always related to the information displayed in the other three panels. The Browse Menu relates to all four panels at the same time.

### ***III. Load multiple text documents into AutoMap***

In order to analyze multiple texts simultaneously, you will need to store the documents together in one folder.

**Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07.**

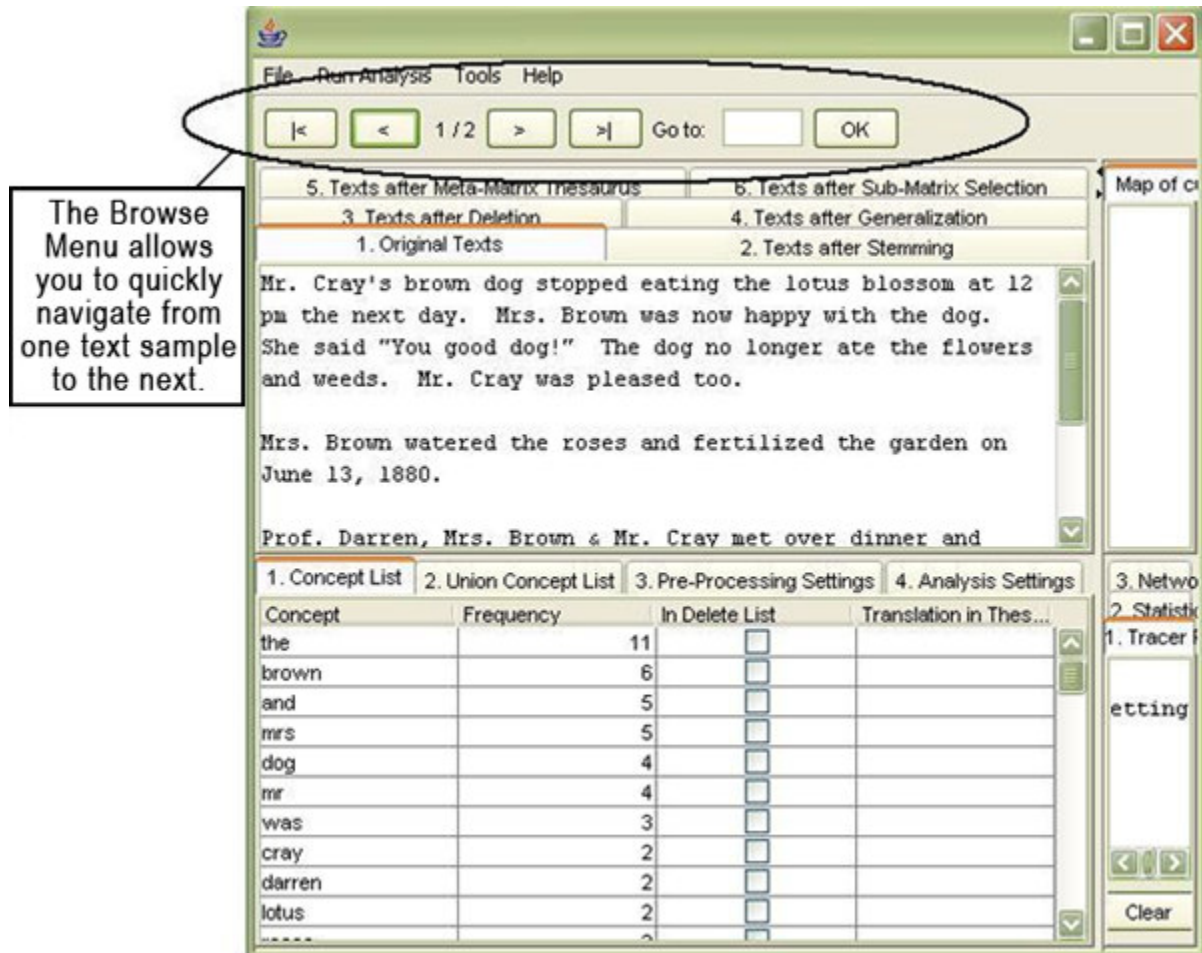
- In the AutoMap menu bar, select File > Open Multiple Files. A folder chooser window will pop up.
- Navigate to the folder containing the documents you wish to analyze, and single-click on it.

## Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07

- Click the Open button in the bottom right corner.

The first text will be displayed in panel P1 under the tab titled 1.Original Texts. You can use the Browse Menu to flip through the documents.

Below is a screen capture of the first text display:



## IV. Carry out a simple content analysis

A simple content analysis determines the frequencies of all words in a text document.

### Examine the Concept List:

The Concept List considers concepts in each text file individually. It is created automatically when a text or set of texts is loaded into AutoMap, and is displayed in panel P2 under the tab titled 1. Concept List.

The Concept List tells you several important facts about each text in your set:

- The number of concepts found in the text displayed in panel P1 under the tab titled 1.Original Text
- Related frequencies of those concepts
- Cumulated unique concepts and total concepts contained in the data set

**Note: The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts.**

By default, the Concept List is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To resort the list by frequency, click on the second-column header Frequency.

Below is a screen capture showing part of the Concept List, sorted by frequency, for the first document in our loaded folder of texts (Text1-in-user's guide.txt):

Click this column heading to sort concepts alphabetically.

Click this column heading to sort concepts by frequency.

Concept	Frequency	In Delete List
the	10	<input type="checkbox"/>
brown	6	<input type="checkbox"/>
dog	6	<input type="checkbox"/>
mrs	5	<input type="checkbox"/>
a	4	<input type="checkbox"/>
and	4	<input type="checkbox"/>
mr	4	<input type="checkbox"/>

## Create and refresh the Union Concept List:

The Union Concept List considers concepts across all texts in a set (unlike the Concept List, which considers each text in the set individually). It is located in panel P2 under the tab titled 2.Union Concept List.

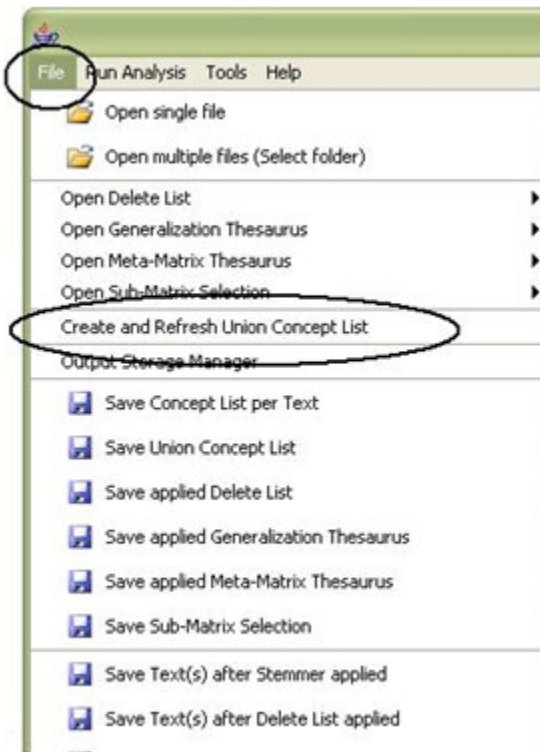
The Union Concept List tells you several important facts about your text set:

- Concepts contained in all loaded text sets
- Related, cumulative frequencies of concepts in all text sets
- Cumulative unique and total concepts

Before viewing your Union Concept List, you must refresh it. From the Menu Bar, choose File > Refresh Union Concept List.

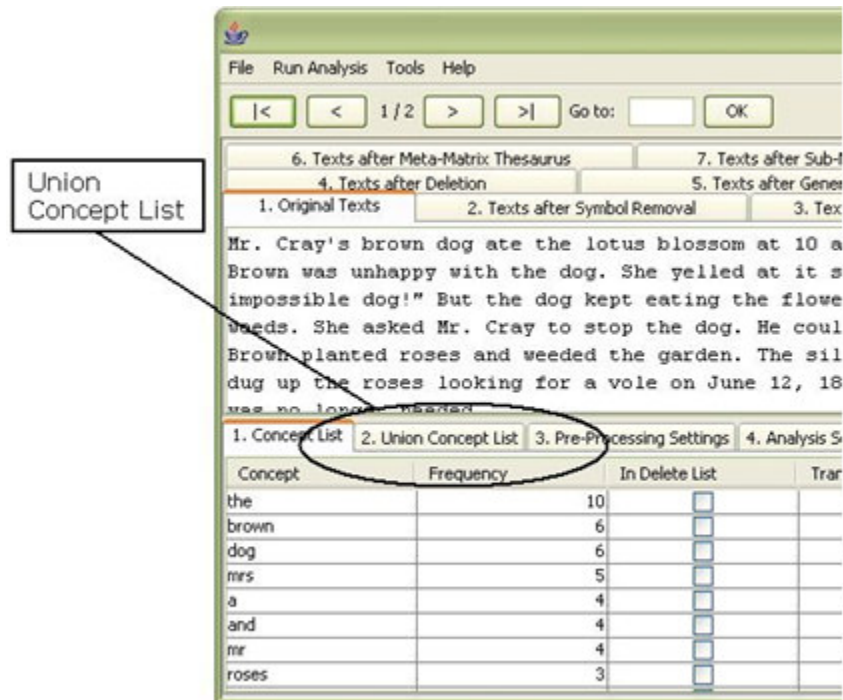
The Union Concept List can be refreshed after each step of pre-processing to visualize the impact of pre-processing operations on the union of concepts.

Below is a screen capture showing where to find the Refresh Union Concept List command:



After refreshing your Union Concept List, you can view it under the 2.Union Concept List tab in panel P2. Below is a screen capture showing where this tab is located in the GUI:





By default, the list is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To re-sort the list by Frequency, click on the second-header column Frequency.

**Note: The number of unique concepts considers each concept only once. The number of total concepts also considers repetitions of concepts.**

## V. The basics of pre-processing in AutoMap

Now you are ready to pre-process the data, or reduce it so that it includes only terms which are relevant to you.

Here are some points to consider:

- Semantic Network Analysis can be run without any prior data pre-processing
- Meta-Matrix Text Analysis and Sub-Matrix Text Analysis require pre-processing

## VI. The basics of delete lists

Deletion removes non-content bearing conjunctions and articles from texts (Carley, 1993). Non-content bearing concepts to be deleted from the texts are denoted in a Delete List. When applying a Delete List, AutoMap searches the text(s) for concepts specified in the Delete List and delete matches from the text(s).



For example:

- Original input text: The New York City Police Department said a number of people were trapped in elevators for awhile
- Entries in the Delete List: the, a, of, were, in, for, awhile
- Text after deletion: New York City Police Department said number people trapped elevators

The Delete List is NOT case sensitive.

You can use the predefined Delete Lists that AutoMap offers, or create your own Delete List.

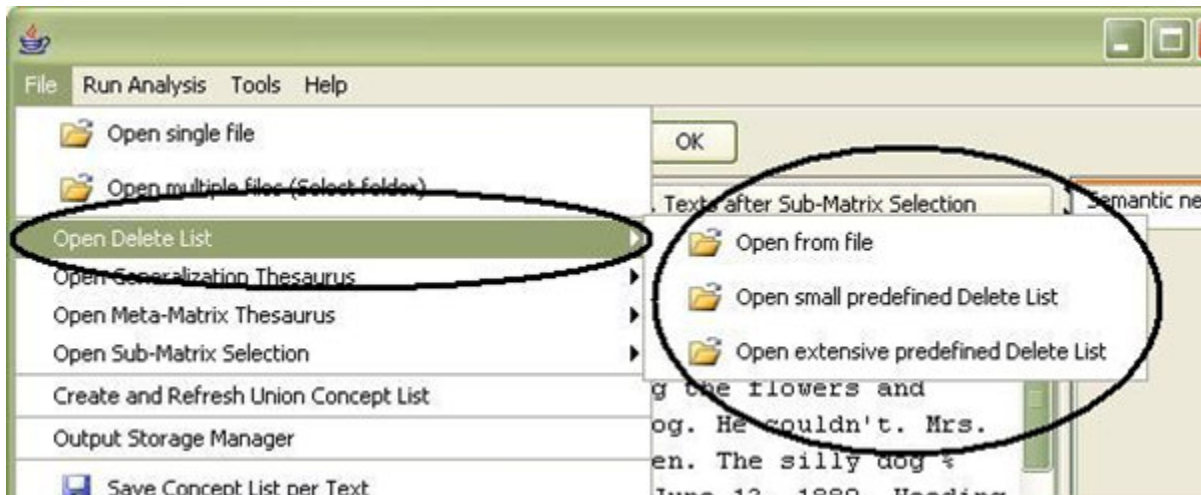
All Delete Lists can be edited.

## ***VII. Open a Delete List***

Select the File menu, then Open Delete List. Choose one of the following options:

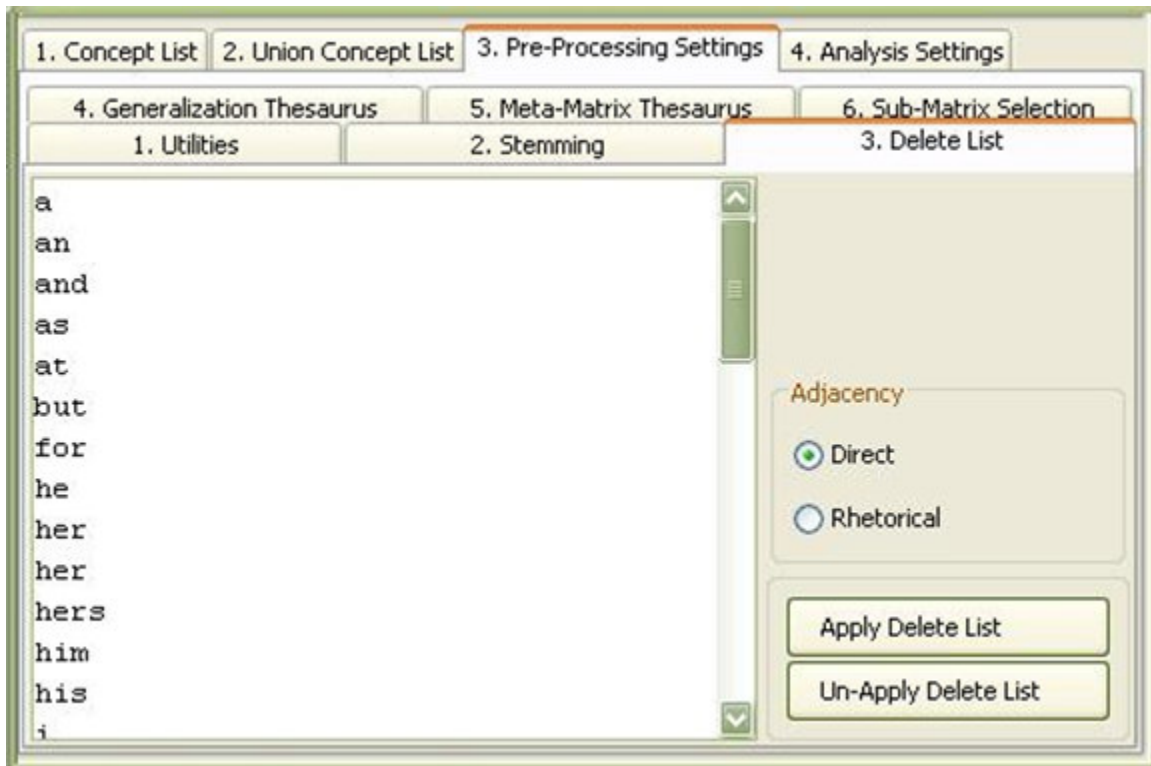
- Open from file: A file chooser appears. Select a delete list and click Open
- Open small predefined Delete List: This opens AutoMap's predefined small delete list
- Open extensive predefined Delete List: This opens AutoMap's extensive small delete list

Below is a screen capture showing where to access the Delete List utility:



The Delete List is located in panel P2 in the AutoMap GUI, under the tab titled 3.Pre-Processing Settings. It appears as a sub-tab titled 3.Delete List.

Below is a screen capture of this location:



The Delete List can be edited by clicking inside the panel where the delete terms are listed in the 3.Delete List index card.

The types of Delete Lists available in AutoMap are:

- Small predefined Delete List  
This is compiled of words that occur most frequently in English: *a, an, and, some, many, this, that, these, those, the, all, one, every.*
- Extensive predefined Delete List  
This is based on words occurring most frequently in English: *a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself.* As the name indicates, the Extensive Delete list contains more words than the Small Predefined Delete List.

### ***VIII. Delete concepts and examine the effects on the text and concept list***

**Note: If you wish to apply a Delete List and a Thesaurus, we recommend applying the Delete List first, followed by the Thesaurus.**

When applying a delete list, AutoMap:

- Searches the text or texts for concepts specified in the Delete List
- Deletes matching concepts from the texts

- Displays the resulting texts in P1 under the 4.Texts After Deletion tab

Follow these steps to apply a delete list:

1. Choose an adjacency option on the 3.Delete List index card. The types of adjacency are the following:
  - Direct Adjacency  
If this option is chosen, concepts in the text that match concepts specified in the delete list are removed from the texts. As a result, concepts on either side of the deleted concept move together and are treated as directly adjacent to each other for further analysis and visualization.
  - Rhetorical Adjacency  
If this option is chosen, concepts in the text that match concepts specified in the delete list are removed from the text, but placeholders ("xxx") are inserted to hold their places. The placeholders retain the original distances of the maintained concepts. This is helpful for visual analysis.

If you do not change the adjacency option, AutoMap uses direct adjacency as the default.

2. Click the Apply Delete List button on the 3.Delete List index card. This deletes the concepts that correspond to concepts specified in the Delete List from all loaded texts.
3. View your newly pre-processed texts in panel P1 under the 4.Texts After Deletion tab.

To apply multiple delete lists, load the first one in and apply it, then load in the next and apply it, and so on.

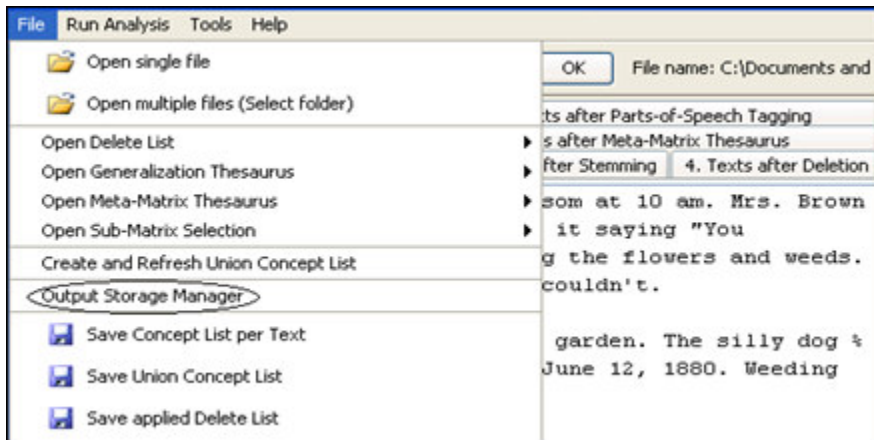
### ***IX. Un-apply a Delete List***

To un-apply a Delete List that was just applied to the data, go to the 3.Delete List index card and click the Un-Apply Delete List button. In panel P2, the 4.Texts After Deletion index card will be cleared.

### ***X. Save an applied Delete List***

To specify a location for your Union Concept List file, choose File in the Menu Bar, then Output Storage Manager.

Below is a screen capture showing where to find this in the GUI:



A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

The Applied Delete List appears in the list on the left-hand side of the Output Storage Manager. Click the Save file as... button to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want and type in a filename.

### **Example: C:\Documents and Settings\carley\Desktop\Lesson 2\SampleDeleteList**

Click the Save Settings button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the red X at the top right corner.

In the Automap Menu Bar, choose File > Save Delete List. This saves the Delete List in the location you specified in the Output Storage Manager.

## ***XI. Save text(s) after application of Delete List***

To determine where your saved texts will end up, choose File > Output Storage Manager (as you did in the previous step).

A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

Find the Texts After Deletion item in the list on the left-hand side of the Output Storage Manager. Click the Save file as... button to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want.

## **Example: C:\Documents and Settings\carley\Desktop\Lesson 2\TextsAfterDeletion**

Click the Save Settings button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the red X at the top right corner.

In the Automap Menu Bar, choose File > Save Text(s) after Delete List Applied. This saves your modified texts in the location you chose in the Output Storage Manager.

### ***XII. Modify a Delete List***

By clicking inside the text field in the 3.Delete List index card, you can:

- Add concepts: Press the <Enter> key after typing in a concept so that there is only one concept per line
- Modify concepts: Go to the desired line and retype the concept
- Drop concepts: Highlight the concept and press the <Delete> key

### ***XIII. Create a delete list from scratch in AutoMap***

There are two ways to create a new delete list:

Within AutoMap

- Go to the 3.Delete List index card and click inside the text field.
- Concepts must be arranged with one per line. To do this, simply press <Enter> after entering each concept.
- Avoid empty lines.
- When you are finished adding concepts, click the Apply Delete List button.

Outside of AutoMap

- Use any text editor to create a Delete List. Keep the following in mind:
- The general structure of a Delete List requires one concept per line.
- Avoid empty lines.
- The Delete List is not case-sensitive.
- Save the Delete List by going to File > Save As, and typing in a filename.

## **Example: C:\Documents and Settings\carley\Desktop\Lesson 2\SampleDeleteList.txt**

- Open your new Delete List in AutoMap: Go to the File > Open Form File, browse to your file, and click Open.

**To delete concepts using this new delete list and examine the effects on the text and concept list, follow Step VIII.**

**To save a delete list you have created from scratch, follow Step X.**

**Example: C:\Documents and Settings\carley\Desktop\Lesson 2\AppliedDeleteList**

#### ***XIV. Open a chosen delete list with Microsoft Excel***

To open your delete list outside of AutoMap, navigate to the location in which the delete list was saved, and double-click on it. A window reading <Open With> pops up. Select Microsoft Excel and click OK. Provided Microsoft Excel is installed on your computer, the file will open.

Alternatively, you can copy and paste your delete list from the AutoMap GUI directly into Microsoft Excel. Press <Control> and <A> simultaneously while in the 3.Delete List panel (this selects all concepts in your delete list), then press <Control> and <C> simultaneously (this copies the delete list). Go to Microsoft Excel and press <Control> and <V> simultaneously (this pastes the delete list into Excel).

#### ***XV. Modify a delete list in Microsoft Excel***

The same rules that applied for delete lists in AutoMap also apply for lists in Excel:

- Use any text editor to create a Delete List. Keep the following in mind:
- The general structure of a Delete List requires one concept per line.
- Avoid empty lines.
- The Delete List is NOT case-sensitive.
- Save the Delete List by going to File > Save As, and typing in a filename.
- Open your new Delete List in AutoMap: Go to the File > Open Form File, browse to your file, and click Open.

To un-apply a Delete List you have created from scratch and/or modified, follow Step IX.  
To load in a modified delete list, follow Step VII.

To delete concepts using a modified delete list and examine the effects on the text and concept list, follow Step VIII.

## Lesson 3

### *Processing data: using a Generalization Thesaurus*

#### *I. Open AutoMap*

##### **On Empire:**

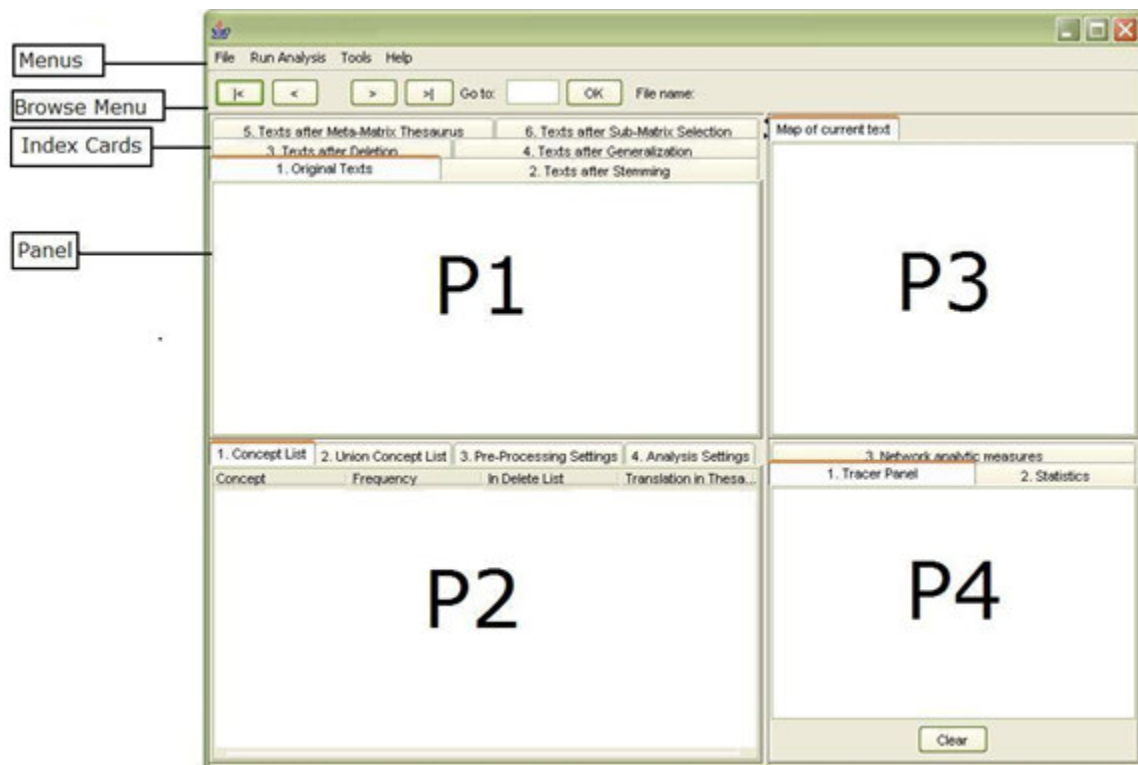
- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click on the Run Automap icon. The Automap Graphical User Interface (GUI) pops up.

From anywhere else: (requires a web connection)

<http://www.casos.cs.cmu.edu/projects/automap/software.html>

#### *II. Overview of the AutoMap Graphical User Interface (GUI)*

Below is a screen capture of the AutoMap GUI:



The AutoMap GUI is divided into four primary quadrants, or panels: P1 (top left), P2 (bottom left), P3 (top right), and P4 (bottom right) as seen in the screen shot above.

A drop-down menu bar at the top of the window provides access to analysis tools and utilities. The Browse Menu allows you to quickly navigate between documents you have loaded into Automap.

The P2 panel can be edited, but the other panels (P1, P3, P4) cannot. Information displayed on each panel is always related to the information displayed in the other three panels. The Browse Menu relates to all four panels at the same time.

### ***III. Load multiple text documents into AutoMap***

In order to analyze multiple documents simultaneously, you will need to store the documents together in one folder.

Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07.

- In the AutoMap menu bar, select File, then Open Multiple Files. A folder chooser window will pop up.
- Navigate to the folder containing the documents you wish to analyze, and single-click on it.

Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07

- Click the Open button in the bottom right corner.

The first text will be displayed in panel P1 under the tab titled 1.Original Texts. You can use the Browse Menu to browse through the texts.

Below is a screen capture of the first text display:

### ***IV. Carry out a simple content analysis***

A simple content analysis determines the frequencies of all words in a text document.

#### **Examine the Concept List:**

The Concept List considers concepts in each text file individually. It is created automatically when a document or set of documents is loaded into Automap, and is displayed in panel P2 under the tab titled 1.Concept List.

The Concept List tells you several important facts about each text in your set:

- The number of concepts found in the text displayed in panel P1 under the tab titled 1.Original Text
- Related frequencies of those concepts
- Cumulated unique concepts and total concepts contained in the data set



**Note: The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts.**

By default, the Concept List is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To return the list to frequency, click on the second-column header Frequency.

Below is a screen capture showing part of the Concept List, sorted by frequency, for the first document in our loaded folder of texts (Text1-in-user's guide.txt):

Create and refresh the Union Concept List:

The Union Concept List considers concepts across all texts in a set (unlike the Concept List, which considers each text in the set individually). It is located in panel P2 under the tab titled 2.Union Concept List.

The Union Concept List tells you several important facts about your text set:

- Concepts contained in all loaded text sets
- Related, cumulative frequencies of concepts in all text sets
- Cumulative unique and total concepts.

Before viewing your Union Concept List, you must refresh it. To refresh, go in the Menu Bar, choose File, then Refresh Union Concept List.

The Union Concept List can be refreshed after each step of pre-processing to visualize the impact of pre-processing operations on the union of concepts.

Below is a screen capture showing where to find the Refresh Union Concept List command:

After refreshing your Union Concept List, you can view it under the 2.Union Concept List tab in panel P2. Below is a screen capture showing where this tab is located in the GUI:

By default, the list is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To return the list by Frequency, click on the second-header column Frequency.

**Note: The number of unique concepts considers each concept only once. The number of total concepts also considers repetitions of concepts.**

## ***V. The basics of pre-processing***

AutoMap follows a hierarchy of pre-processing techniques. If you apply a pre-processing technique of a lower order prior to a technique of higher order, the pre-processing will be maintained through all following procedures of higher order. If needed, you can un-apply each technique after applying it.

If you wish to apply multiple pre-processing techniques, carry out the process in the following order:

1. Named-Entity Recognition  
This utility does not impact the data. It can be used before any type of analysis is run. It can be used before or after Stemming.
2. Collocation/Bigram Identification  
This utility does not impact the data. It can be used before any type of analysis is run.
3. Stemming  
This can be used before any type of analysis is run. It can be used before or after Named-Entity Recognition.
4. Deletion  
This can be used before any type of analysis is run.
5. Thesauri
  - Generalization Thesaurus: This can be applied before Semantic Network Analysis is run. It can be applied before Meta-Matrix Thesaurus is applied.
  - Meta-Matrix Thesaurus: This has to be applied if a Meta-Matrix Analysis will be run.
  - Sub-Matrix Selection: This can only be performed if the Meta-Matrix Thesaurus was applied. It must be applied if a Sub-Matrix Analysis will be run.

This hierarchy is reflected in the numbering of the index card tabs in panels P1 and P2, to make the sequence of the pre-processing steps more intuitive.

Below is a screen capture showing the hierarchy of pre-processing techniques in panels P1 (top) and P2 (bottom):

## ***VI. Open a Generalization Thesaurus***

From the Menu bar select File > Open Generalization Thesaurus.

The thesaurus will be displayed in panel P2 under the 3.Preprocessing Settings tab, as a sub-tab titled 4.Generalization Thesaurus.

## ***VII. Apply a thesaurus and examine its effect on the texts and concept list***

**Note: If you wish to apply both a Delete List and a Generalization Thesaurus, be sure to use first the Delete List and then the Thesaurus.**

When applying a thesaurus, you have the following options:

### **Thesaurus content only**

If this option is chosen, AutoMap:

1. Searches the texts for concepts specified in the thesaurus.
2. Translates matches into key concepts.
3. Maintains only key concepts in the pre-processed texts. The remaining input text is dropped and is not considered in further pre-processing or analysis. The original distances of the key concepts are not maintained, but punctuation marks (such as at the ends of sentences and paragraphs) are maintained and considered in analysis.
4. Results in all key concepts in the text appearing directly adjacent to each other.

If this option is not chosen, Automap:

1. Searches the texts for concepts specified in the thesaurus.
2. Translates matches into key concepts.
3. Keeps the rest of the text as is. This means that all other concepts in the text that did not match concepts specified in the thesaurus will not be affected in any way. Original distances of both unaffected concepts and key concepts are maintained (unless a concept consisting of more than one word was translated into a key concept).

If you do not choose “thesaurus content only”, the setting will not be applied. If you do choose “thesaurus content only”, you then have another option:

### **Direct or rhetorical adjacency**

Direct adjacency means that original distances between concepts that represent the key concepts are neither visualized nor considered for analysis.

Rhetorical adjacency means that the original distances between key concepts are retained and incorporated into later analyses. The original distances are visually symbolized by placeholders (“xxx”).

To choose the direct adjacency option, click the Direct button in the Adjacency field in the 4.Generalization Thesaurus index card under 3.Pre-processing tab in panel P2. To choose the rhetorical adjacency option, click the Rhetorical button in the same field.

If you do not change the adjacency option, AutoMap uses direct adjacency as the default. Also, direct adjacency is automatically applied if the “thesaurus content only” option is not checked.

To select the thesaurus content only option, check the box for Thesaurus content only the 4.Generalization Thesaurus index card in panel P2. You can then choose either direct or rhetorical adjacency by clicking in the circles with those labels.

To apply your chosen settings, click the Apply button on the 4.Generalization Thesaurus index card in panel P2.

To switch back to not using the “thesaurus content only” option, un-check the Thesaurus content only box in the 4.Generalization Thesaurus index card in panel P2, then apply the Generalization Thesaurus again.

### ***VIII. Un-apply a Generalization Thesaurus***

To un-apply a Generalization Thesaurus that was applied to the data, go to the 4.Generalization Thesaurus index card and click the Un-Apply button. This clears the 5.Texts After Generalization tab in panel P1.

### ***IX. Modify the generalization thesaurus list***

You can modify your generalization thesaurus by clicking inside the text field on the 4.Generalization Thesaurus index card in panel P2. Keep in mind the following points:

1. Every line contains Concept / Key Concept (in other words, Old Word / New Word.)
2. A Concept can be one or more words.
3. A Key Concept is one word.
4. Be sure to separate the words by columns.
5. The Thesaurus is not case sensitive.

To apply a modified thesaurus and examine its effects on the texts and concept list, follow Step VII.

To un-apply a modified thesaurus, follow Step VIII.

### ***X. Create a generalization thesaurus list from scratch***

You can create a Generalization Thesaurus from scratch in AutoMap.

Find the 4.Generalization Thesaurus index card under the 3.Preprocessing Settings tab in panel P2. You will see a text field, composed of two columns titled “concept” and “generalization”, that resembles a Microsoft Excel spreadsheet. To edit items inside this text field, simply double-click on them.

Automap can help with the building of a generalization thesaurus by loading the Union Concept List from the highest level of pre-processing applied into the Generalization Thesaurus field.

To load in the highest-level union concept list:

1. Follow Step 4 to create and refresh the Union Concept List
2. Click the Load Union Concept List button on the 4.Generalization Thesaurus index card.

Once loaded into AutoMap, the Concept List can be refined by applying Named-Entity Recognition and Deletion prior to application of the Generalization Thesaurus. (See Lessons 1 and 2).

To apply the thesaurus you have made from scratch and examine its effect on the texts and concept list, follow Step VII.

## ***XI. Save the Generalization Thesaurus***

To specify a location for your Generalization Thesaurus file, select File in the Menu Bar, then Output Storage Manager.

Below is a screen capture showing where to find this in the GUI:

A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

Find the Applied Generalization Thesaurus item in the list on the left-hand side of the Output Storage Manager. Click the Save file as... button to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want and type in a filename.

### **Example: C:\Documents and Settings\carley\Desktop\Lesson 3\AppliedGeneralizationThesaurus**

Click the Save Settings button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the red X at the top right corner.

In the Automap Menu Bar, choose File, then Save applied Generalization Thesaurus.

## ***XII. Open the Generalization Thesaurus in Microsoft Excel***

Outside of Automap, navigate to the Generalization Thesaurus you have saved and double-click on it. A window titled “Open With” will pop up. Double-click on Microsoft Excel. The file will open in Excel (provided you have Excel installed on your computer).

Alternatively, you can copy and paste your Generalization Thesaurus into Excel directly from Automap. Single-click inside the text field on the 4.Generalization Thesaurus index card in panel P2. Press <Control> and <A>; simultaneously to select all the text, followed by and simultaneously to copy the selected text. Then go into a blank Excel spreadsheet and press and simultaneously to paste the generalization thesaurus into the spreadsheet.

## ***XII. Modify the Generalization Thesaurus in Microsoft Excel***

You can modify your generalization thesaurus in Microsoft Excel just as for a standard spreadsheet. Keep in mind the following points:

1. Every line contains Concept / Key Concept (in other words, Old Word / New Word.)
2. A Concept can be one or more words.
3. A Key Concept is one word.
4. Be sure to separate the words by columns.
5. The Thesaurus is not case sensitive.

To load in a thesaurus you have modified in Microsoft Excel, follow Step VI.

To apply this modified thesaurus and examine its effects on the texts and concept list, follow Step VII.

To un-apply this modified thesaurus, follow Step VIII.

## **Lesson 4**

### ***AutoMap Usage "Snapshots"***

#### ***1. Snapshot: Split Input Text Files***

##### *Purpose*

Split large text files into smaller ones of minimum equal size.

When to apply it: In order to speed up AutoMap coding.

Input from user: Number of words (NW) that each text file should contain after splitting.

### *How it works*

Each text will be split at the next sentence mark after the number of words that the user had specified. Thus, each resulting split text will contain at least NW words.

### *Output*

N texts that contain at least NW in directory specified by the user. The resulting texts maintain the original filename plus a counter, starting from 0 and going up to N, where N indicates the largest number of texts that an original text had been split up into.

### *How To*

Click the Tools menu and select Open Text File Splitter. Follow the directions specified in the user interface.

## **2. Snapshot: Using Compare Maps**

### *Purpose*

- Compare individual mental maps and structures of social and organizational systems extracted from texts with AutoMap by using set theory.
- Consolidate individual mental maps and structures of social and organizational systems extracted from texts with AutoMap to a degree specified by the user.

### *How To*

Click the Tools menu and select Open CompareMap. For further instructions consult the CompareMap User's Guide.

## **3. Snapshot: Merge DyNetML Files**

### *Purpose*

Merge multiple DyNetML files into 1 DyNetML file.

### *Example*

This might be needed for example when DyNetML files that was generated per text during Sub-Matrix Text Analysis need to be consolidated into 1 DyNetML file that represent the entire text set.

### *Output*

1 DyNetML file.

### *How To*

Click the Tools menu and select Open DyNetML File Merger. Follow the directions specified in the user interface.

## **4. Snapshot: Convert Network Data Formats**

### *Purpose*

Convert a file in a specific network data format (CVS, DL, UCINET, DyNetML, VNA) into another network data format.

### *How To*

Click the Tools menu and select Open Matrix Editor. Follow the directions specified in the user interface.

## **5. Snapshot: Edit Network Data**

### *Purpose*

Edit relational data.

### *How To*

Click the Tools menu and select Open Network Data Format Converter. Follow the directions specified in the user interface.

## **6. Snapshot: Visualize Semantic Networks**

### *Purpose*

Visualize mental models and social structure.

### *How To*

- Within AutoMap: click the Tools menu and select Open SocialInsight Visualizer

*Example: Load in DyNetML files created in AutoMap.*

- With External Tools: Convert DyNetML or DL files generated with AutoMap in formats required by the external tool you want to use:

*Example: Convert DL into VNA files can be visualized in NetDraw. In NetDraw open VNA file: File > Open > VNA text file > complete.*



## References

- Borgatti, S.P., Everett, M.G., & Freeman, L.C. (ac) (2002).* UCINET. for Windows. Software for Social Network Analysis. Analytic Technologies, Inc.
- Burkart, M. (1997).* Thesaurus. In M. Buder, W. Rehfeld, T. Seeger & D. Strauch (Eds.), Grundlagen der praktischen Information und Dokumentation: Ein Handbuch zur Einführung in die fachliche Informationsarbeit. (pp. 160 - 179). 4th edition. München: Saur.
- Kathleen M. Carley, 1997, "Extracting Team Mental Models Through Textual Analysis."* Journal of Organizational Behavior, 18: 533-538.
- Kathleen M. Carley, 1994, "Extracting Culture through Textual Analysis."* Poetics, 22: 291-312.
- Kathleen M. Carley and David Kaufer, 1993, "Semantic Connectivity: An Approach for Analyzing Semantic Networks."* Communication Theory, 3(3): 183-213.
- Kathleen M. Carley, 1993, "Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis."* In Marsden P. (Ed), Sociological Methodology, 23: 75-126. Oxford: Blackwell.
- Kathleen M. Carley, 1997, "Network Text Analysis: The Network Position of Concepts."* Chapter 4 in C. Roberts (Ed.), Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 79-100.
- Carley, K.M. & Reminga, J. (2004).* ORA: Organization Risk Analyzer. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report.
- Carley, K.M. (2003).* Dynamic Network Analysis. In R. Breiger, K.M. Carley & P. Pattison (Eds.), Summary of the NRC workshop on social network modeling and analysis, (pp. 133-145). Committee on Human Factors, National Research Council.
- Kathleen M. Carley and Michael Palmquist, 1992, "Extracting, Representing and Analyzing Mental Models."* Social Forces, 70(3): 601-636
- Kathleen M. Carley, 1993, "Content Analysis,"* in Asher R.E. et al. (Eds.), The Encyclopedia of Language and Linguistics. Edinburgh, UK: Pergamon Press. Vol. 2: 725-730.
- Diesner, J., & Carley, K.M., 2005, "Exploration of Communication Networks from the Enron Email Corpus,"* Proceedings of the Workshop on Link Analysis, Counterterrorism

and Security, SIAM International Conference on Data Mining 2005, pp. 3-14. Newport Beach, CA, April 21-23, 2005.

*Jana Diesner, Kathleen M. Carley, 2005, "Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis."* In V.K. Narayanan & D.J. Armstrong (Eds.) *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations*, Chapter 4, Harrisburg, PA: Idea Group Publishing.

*Diesner, J., & Carley, K.M. (2004), AutoMap1.2 - Extract, analyze, represent, and compare mental models from texts.* Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-100. URL: <http://reports-archive.adm.cs.cmu.edu/isri2004.html> (01-22-2004).

*Diesner, J., & Carley, K.M. (2005). Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis.* Chapter 4 in V.K. Narayanan, & D.J. Armstrong (Eds.), *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations* (pp.81-108). Harrisburg, PA: Idea Group Publishing.

*Jurafsky, D., & Marton, J.H. (2000). Speech and Language Processing.* Upper Saddle River, New Jersey: Prentice Hall.

*David Kaufer and Kathleen M. Carley, 1993, "Condensation Symbols: Their Variety and Rhetorical Function in Political Discourse."* *Philosophy and Rhetoric*, 26(3): 201-226.

*Klein, H. (1997), Classification of Text Analysis Software.* In R. Klar, & O. Opitz (Eds.), *Classification and knowledge organization: Proceedings of the 20th annual conference of the Gesellschaft für Klassifikation e.V.*, (pp. 255-261). University of Freiburg , 1996, Berlin , New York : Springer. 1997.

*Krovetz, Robert (1995). Word Sense Disambiguation for Large Text Databases.* Unpublished PhD Theis, University of Massachusetts.

*Magnini, B., Negri, M., Prevete, R., & Tanev, H. (2002). A WordNet-based approach to Named Entities Recognition. In Proceedings of SemaNet'02: Building and Using Semantic Networks.* (pp. 38-44). Taipei, Taiwan, August 2002.

*Mrvar, A. Centrality measures.* URL: <http://mrvar.fdv.uni-lj.si/sola/info4/uvod/part4.pdf> ( 06-13-2004 )

*Michael Palmquist, Kathleen M. Carley, and Thomas Dale, 1997, "Two applications of automated text analysis: Analyzing literary and non-literary texts."* Chapter 10 in C. Roberts (Ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 171-189.

*Popping, R., & Roberts, C.W. (1997). Network Approaches in Text Analysis. In R. Klar & O. Opitz (Eds.), Classification and Knowledge Organization: Proceedings of the 20th annual conference of the Gesellschaft für Klassifikation e.V., (pp. 381-898). University of Freiburg , Berlin , New York : Springer.*

*Porter, M.F. 1980. An algorithm for suffix stripping. I 14 (3): 130-137.*

*Tsvetovat, M., Reminga, J., & Carley, K.M. (2004). DyNetML: Interchange Format for Rich Social Network Data. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, CASOS Technical Report CMU-ISRI-04-105. URL: <http://reports-archive.adm.cs.cmu.edu/anon/isri2004/abstracts/04-105.html> .*

*Zuell, C., & Alexa, M. (2001). Automatisches Codieren von Textdaten. Ein Ueberblick ueber neue Entwicklungen. In W. Wirth & E. Lauf (Eds.), Inhaltsanalyse - Perspektiven, Probleme, Potenziale (pp. 303-317). Koeln: Herbert von Halem.*

## ***Where to learn more about Dynamic Network Analysis***

*Kathleen M. Carley, 2003, “Dynamic Network Analysis” in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Ronald Breiger, Kathleen Carley, and Philippa Pattison, (Eds.) Committee on Human Factors, National Research Council, National Research Council. Pp. 133-145, Washington, DC.*

*Kathleen M. Carley, 2002, “Smart Agents and Organizations of the Future“ The Handbook of New Media. Edited by Leah Lievrouw and Sonia Livingstone, Ch. 12, pp. 206-220, Thousand Oaks, CA, Sage.*

*Kathleen M. Carley, Jana Diesner, Jeffrey Reminga, Maksim Tsvetovat, 2005-forthcoming, Toward an Interoperable Dynamic Network Analysis Toolkit, DSS Special Issue on Cyberinfrastructure for Homeland Security: Advances in Information Sharing, Data Mining, and Collaboration Systems.*

## **Entropy references:**

### **Theory:**

*W. Weaver & C. E. Shannon, The Mathematical Theory of Communication, Urbana, Illinois: University of Illinois Press, 1949*

### **Empiric application:**

*Diesner, J., Carley, K.M., & Katzmaier, H. (2007). The morphology of a breakdown. How the semantics and mechanics of communication networks from an organization in crises relate. XXVII Sunbelt Social Network Conference, Corfu, Greece, May 2007.*